

21st Edition

• **HARRISON'S**
PRINCIPLES OF
**INTERNAL
MEDICINE**

LOSCALZO

FAUCI

KASPER

HAUSER

LONGO

JAMESON

VOLUME 1

Mc
Graw
Hill



21st Edition

HARRISON'S®

PRINCIPLES OF

**INTERNAL
MEDICINE**

Editors of Previous Editions

T. R. Harrison

Editor-in-Chief, Editions 1, 2, 3, 4, 5

W. R. Resnick

Editor, Editions 1, 2, 3, 4, 5

M. M. Wintrobe

Editor, Editions 1, 2, 3, 4, 5

Editor-in-Chief, Editions 6, 7

G. W. Thorn

Editor, Editions 1, 2, 3, 4, 5, 6, 7

Editor-in-Chief, Edition 8

R. D. Adams

Editor, Editions 2, 3, 4, 5, 6, 7, 8, 9, 10

P. B. Beeson

Editor, Editions 1, 2

I. L. Bennett, Jr.

Editor, Editions 3, 4, 5, 6

E. Braunwald

Editor, Editions 6, 7, 8, 9, 10, 12, 13, 14, 16, 17

Editor-in-Chief, Editions 11, 15

K. J. Isselbacher

Editor, Editions 6, 7, 8, 10, 11, 12, 14

Editor-in-Chief, Editions 9, 13

R. G. Petersdorf

Editor, Editions 6, 7, 8, 9, 11, 12

Editor-in-Chief, Edition 10

J. D. Wilson

Editor, Editions 9, 10, 11, 13, 14

Editor-in-Chief, Edition 12

J. B. Martin

Editor, Editions 10, 11, 12, 13, 14

A. S. Fauci

Editor, Editions 11, 12, 13, 15, 16, 18, 19, 20, 21

Editor-in-Chief, Editions 14, 17

R. Root

Editor, Edition 12

D. L. Kasper

Editor, Editions 13, 14, 15, 17, 18, 20, 21

Editor-in-Chief, Editions 16, 19

S. L. Hauser

Editor, Editions 14, 15, 16, 17, 18, 19, 20, 21

D. L. Longo

Editor, Editions 14, 15, 16, 17, 19, 20, 21

Editor-in-Chief, Edition 18

J. L. Jameson

Editor, Editions 15, 16, 17, 18, 19, 21

Editor-in-Chief, Edition 20

J. Loscalzo

Editor, Editions 17, 18, 19, 20

Editor-in-Chief, Edition 21



21st Edition

HARRISON'S®

PRINCIPLES OF

**INTERNAL
MEDICINE**

Editors

Joseph Loscalzo, MD, PhD

Hersey Professor of the Theory and Practice of Medicine, Harvard Medical School; Chairman, Department of Medicine; Soma Weiss MD Distinguished Chair in Medicine; Physician-in-Chief, Brigham and Women's Hospital, Boston, Massachusetts

Dennis L. Kasper, MD

William Ellery Channing Professor of Medicine and Professor of Immunology, Department of Immunology, Harvard Medical School; Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts

Dan L. Longo, MD

Professor of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital; Deputy Editor, *New England Journal of Medicine*, Boston, Massachusetts

Anthony S. Fauci, MD

Chief, Laboratory of Immunoregulation; Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland

Stephen L. Hauser, MD

Robert A. Fishman Distinguished Professor, Department of Neurology; Director, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California

J. Larry Jameson, MD, PhD

Robert G. Dunlop Professor of Medicine; Dean, Raymond and Ruth Perelman School of Medicine; Executive Vice President, University of Pennsylvania for the Health System, Philadelphia, Pennsylvania

VOLUME I



New York Chicago San Francisco Athens London Madrid Mexico City
New Delhi Milan Singapore Sydney Toronto

Copyright © 2022 by McGraw Hill LLC. Previous editions copyright © 2018, 2015 by McGraw-Hill Education; 2012, 2008, 2005, 2001, 1998, 1994, 1991, 1987, 1983, 1980, 1977, 1974, 1970, 1966, 1962, 1958 by The McGraw-Hill Companies, Inc. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-1-26-426851-1
MHID: 1-26-426851-3

The material in this eBook also appears in the print version of this title: ISBN: 978-1-26-426850-4,
MHID: 1-26-426850-5.

eBook conversion by codeMantra
Version 1.0

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill Education eBooks are available at special quantity discounts to use as premiums and sales promotions or for use in corporate training programs. To contact a representative, please visit the Contact Us page at www.mhprofessional.com.

TERMS OF USE

This is a copyrighted work and McGraw-Hill Education and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill Education's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL EDUCATION AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill Education and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill Education nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill Education has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill Education and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

NOTICE

Medicine is an ever-changing science. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy are required. The authors and the publisher of this work have checked with sources believed to be reliable in their efforts to provide information that is complete and generally in accord with the standards accepted at the time of publication. However, in view of the possibility of human error or changes in medical sciences, neither the authors nor the publisher nor any other party who has been involved in the preparation or publication of this work warrants that the information contained herein is in every respect accurate or complete, and they disclaim all responsibility for any errors or omissions or for the results obtained from use of the information contained in this work. Readers are encouraged to confirm the information contained herein with other sources. For example and in particular, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this work is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs.

Cover Illustration

Beginning with the 6th edition, the cover of *Harrison's* has included an image of a bright light—a patient's perception of being examined with an ophthalmoscope. This allegorical symbol of *Harrison's* is a reminder of how the light of knowledge empowers physicians to better diagnose and treat diseases that ultimately afflict all of humankind.

Author Disclosure Policy: McGraw Hill and the *Harrison's* Editorial Board require all contributors to disclose to the Editors and the Publisher any potential financial or professional conflicts that would raise the possibility of distorting the preparation of a *Harrison's* chapter.

Contents

Contributors	xviii
Preface	xxxix
Harrison's Related Resources	xl

PART 1 The Profession of Medicine

1 The Practice of Medicine	1
<i>The Editors</i>	
2 Promoting Good Health	8
<i>Donald M. Lloyd-Jones, Kathleen M. McKibbin</i>	
3 Vaccine Opposition and Hesitancy	13
<i>Julie A. Bettinger, Hana Mitchell</i>	
4 Decision-Making in Clinical Medicine.....	21
<i>Daniel B. Mark, John B. Wong</i>	
5 Precision Medicine and Clinical Care	30
<i>The Editors</i>	
6 Screening and Prevention of Disease	37
<i>Katrina A. Armstrong, Gary J. Martin</i>	
7 Global Diversity of Health System Financing and Delivery	42
<i>Richard B. Saltman</i>	
8 The Safety and Quality of Health Care.....	50
<i>David W. Bates</i>	
9 Diagnosis: Reducing Errors and Improving Quality.....	54
<i>Gordon Schiff</i>	
10 Racial and Ethnic Disparities in Health Care.....	59
<i>Lenny López, Joseph R. Betancourt</i>	
11 Ethical Issues in Clinical Medicine.....	67
<i>Christine Grady, Bernard Lo</i>	
12 Palliative and End-of-Life Care	72
<i>Ezekiel J. Emanuel</i>	

PART 2 Cardinal Manifestations and Presentation of Diseases

SECTION 1 Pain

13 Pain: Pathophysiology and Management	91
<i>James P. Rathmell, Howard L. Fields</i>	
14 Chest Discomfort	100
<i>David A. Morrow</i>	
15 Abdominal Pain.....	108
<i>Danny O. Jacobs</i>	
16 Headache.....	112
<i>Peter J. Goadsby</i>	
17 Back and Neck Pain	117
<i>John W. Engstrom</i>	

SECTION 2 Alterations in Body Temperature

18 Fever	130
<i>Neeraj K. Surana, Charles A. Dinarello, Reuven Porat</i>	
19 Fever and Rash	133
<i>Elaine T. Kaye, Kenneth M. Kaye</i>	
20 Fever of Unknown Origin	145
<i>Chantal P. Bleeker-Rovers, Catharina M. Mulders-Manders, Jos W. M. van der Meer</i>	

SECTION 3 Nervous System Dysfunction

21 Syncope	152
<i>Roy Freeman</i>	
22 Dizziness and Vertigo.....	159
<i>Mark F. Walker, Robert B. Daroff</i>	
23 Fatigue	162
<i>Jeffrey M. Gelfand, Vanja C. Douglas</i>	
24 Neurologic Causes of Weakness and Paralysis.....	165
<i>Stephen L. Hauser</i>	
25 Numbness, Tingling, and Sensory Loss	168
<i>Stephen L. Hauser</i>	
26 Gait Disorders, Imbalance, and Falls	173
<i>Jessica M. Baker</i>	
27 Confusion and Delirium	178
<i>S. Andrew Josephson, Bruce L. Miller</i>	
28 Coma.....	183
<i>S. Andrew Josephson, Allan H. Ropper, Stephen L. Hauser</i>	
29 Dementia.....	189
<i>William W. Seeley, Gil D. Rabinovici, Bruce L. Miller</i>	
30 Aphasia, Memory Loss, and Other Cognitive Disorders.....	195
<i>M.-Marsel Mesulam</i>	
31 Sleep Disorders.....	204
<i>Thomas E. Scammell, Clifford B. Saper, Charles A. Czeisler</i>	

SECTION 4 Disorders of Eyes, Ears, Nose, and Throat

32 Disorders of the Eye.....	215
<i>Jonathan C. Horton</i>	
33 Disorders of Smell and Taste	232
<i>Richard L. Doty, Steven M. Bromley</i>	
34 Disorders of Hearing.....	238
<i>Anil K. Lalwani</i>	
35 Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat.....	247
<i>Rachel L. Amdur, Jeffrey A. Linder</i>	
36 Oral Manifestations of Disease	256
<i>Samuel C. Durso</i>	

SECTION 5 Alterations in Circulatory and Respiratory Functions

37 Dyspnea	263
<i>Rebecca M. Baron</i>	
38 Cough	267
<i>Christopher H. Fanta</i>	
39 Hemoptysis	270
<i>Carolyn M. D'Ambrosio</i>	
40 Hypoxia and Cyanosis.....	272
<i>Joseph Loscalzo</i>	
41 Edema	275
<i>Joseph Loscalzo</i>	
42 Approach to the Patient with a Heart Murmur	278
<i>Patrick T. O'Gara, Joseph Loscalzo</i>	
43 Palpitations.....	286
<i>Joseph Loscalzo</i>	

SECTION 6 Alterations in Gastrointestinal Function

44 Dysphagia.....	287
<i>Ikuko Hirano, Peter J. Kahrilas</i>	
45 Nausea, Vomiting, and Indigestion	291
<i>William L. Hasler</i>	
46 Diarrhea and Constipation.....	297
<i>Michael Camilleri, Joseph A. Murray</i>	
47 Unintentional Weight Loss	309
<i>J. Larry Jameson</i>	
48 Gastrointestinal Bleeding.....	311
<i>Loren Laine</i>	
49 Jaundice.....	315
<i>Savio John, Daniel S. Pratt</i>	
50 Abdominal Swelling and Ascites	321
<i>Lawrence S. Friedman</i>	

SECTION 7 Alterations in Renal and Urinary Tract Function

51 Interstitial Cystitis/Bladder Pain Syndrome.....	325
<i>R. Christopher Doiron, J. Curtis Nickel</i>	
52 Azotemia and Urinary Abnormalities	331
<i>David B. Mount</i>	
53 Fluid and Electrolyte Disturbances	338
<i>David B. Mount</i>	
54 Hypercalcemia and Hypocalcemia.....	356
<i>Sundeep Khosla</i>	
55 Acidosis and Alkalosis.....	359
<i>Thomas D. DuBose, Jr.</i>	

SECTION 8 Alterations in the Skin

56 Approach to the Patient with a Skin Disorder	369
<i>Kim B. Yancey, Thomas J. Lawley</i>	
57 Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders.....	374
<i>Leslie P. Lawley, Justin T. Cheeley, Robert A. Swerlick</i>	
58 Skin Manifestations of Internal Disease	383
<i>Jean L. Bolognia, Jonathan S. Leventhal, Irwin M. Braverman</i>	
59 Immunologically Mediated Skin Diseases	400
<i>Kim B. Yancey, Benjamin F. Chong, Thomas J. Lawley</i>	
60 Cutaneous Drug Reactions	407
<i>Robert G. Micheletti, Misha Rosenbach, Bruce U. Wintrob, Kanade Shinkai</i>	
61 Photosensitivity and Other Reactions to Sunlight.....	417
<i>Alexander G. Marneros, David R. Bickers</i>	

SECTION 9 Hematologic Alterations

62 Interpreting Peripheral Blood Smears.....	424
<i>Dan L. Longo</i>	
63 Anemia and Polycythemia	431
<i>John W. Adamson, Dan L. Longo</i>	
64 Disorders of Granulocytes and Monocytes	439
<i>Steven M. Holland, John I. Gallin</i>	
65 Bleeding and Thrombosis.....	450
<i>Barbara A. Konkle</i>	
66 Enlargement of Lymph Nodes and Spleen	457
<i>Dan L. Longo</i>	

PART 3 Pharmacology

67 Principles of Clinical Pharmacology	465
<i>Dan M. Roden</i>	
68 Pharmacogenomics.....	474
<i>Dan M. Roden</i>	

PART 4 Oncology and Hematology**SECTION 1 Neoplastic Disorders**

69 Approach to the Patient with Cancer	481
<i>Dan L. Longo</i>	
70 Prevention and Early Detection of Cancer.....	490
<i>Jennifer M. Croswell, Otis W. Brawley, Barnett S. Kramer</i>	
71 Cancer Genetics.....	498
<i>Fred Bunz, Bert Vogelstein</i>	
72 Cancer Cell Biology.....	508
<i>Jeffrey W. Clark, Dan L. Longo</i>	
73 Principles of Cancer Treatment	529
<i>Edward A. Sausville, Dan L. Longo</i>	
74 Infections in Patients with Cancer	556
<i>Robert W. Finberg</i>	
75 Oncologic Emergencies.....	565
<i>Rasim Guçalp, Janice P. Dutcher</i>	
76 Cancer of the Skin.....	578
<i>Brendan D. Curti, John T. Vetto, Sancy A. Leachman</i>	
77 Head and Neck Cancer	590
<i>Everett E. Vokes</i>	
78 Neoplasms of the Lung	594
<i>Leora Horn, Wade T. Iams</i>	
79 Breast Cancer	611
<i>Daniel F. Hayes, Marc E. Lippman</i>	
80 Upper Gastrointestinal Tract Cancers.....	626
<i>David Kelsen</i>	
81 Lower Gastrointestinal Cancers	636
<i>Robert J. Mayer</i>	
82 Tumors of the Liver and Biliary Tree	643
<i>Josep M. Llovet</i>	
83 Pancreatic Cancer	657
<i>Daniel D. Von Hoff</i>	
84 Gastrointestinal Neuroendocrine Tumors.....	663
<i>Matthew H. Kulke</i>	
85 Renal Cell Carcinoma	673
<i>Robert J. Motzer, Martin H. Voss</i>	
86 Cancer of the Bladder and Urinary Tract.....	676
<i>Noah M. Hahn</i>	
87 Benign and Malignant Diseases of the Prostate.....	681
<i>Howard I. Scher, James A. Eastham</i>	
88 Testicular Cancer	689
<i>David J. Vaughn</i>	
89 Gynecologic Malignancies	695
<i>David Spriggs</i>	
90 Primary and Metastatic Tumors of the Nervous System	701
<i>Lisa M. DeAngelis, Patrick Y. Wen</i>	
91 Soft Tissue and Bone Sarcomas and Bone Metastases	712
<i>Shreyaskumar R. Patel</i>	

92 Carcinoma of Unknown Primary	716
<i>Kanwal Raghav, James L. Abbruzzese, Gauri R. Varadhachary</i>	
93 Paraneoplastic Syndromes: Endocrinologic/ Hematologic	721
<i>J. Larry Jameson, Dan L. Longo</i>	
94 Paraneoplastic Neurologic Syndromes and Autoimmune Encephalitis.....	728
<i>Josep Dalmau, Myrna R. Rosenfeld, Francesc Graus</i>	
95 Cancer Survivorship and the Long-Term Impact of Cancer and Its Treatment	736
<i>Mark Roschewski, Dan L. Longo</i>	

SECTION 2 Hematopoietic Disorders

96 Hematopoietic Stem Cells.....	743
<i>David T. Scadden, Dan L. Longo</i>	
97 Iron Deficiency and Other Hypoproliferative Anemias	747
<i>John W. Adamson</i>	
98 Disorders of Hemoglobin.....	754
<i>Martin H. Steinberg</i>	
99 Megaloblastic Anemias	766
<i>A. Victor Hoffbrand</i>	
100 Hemolytic Anemias	776
<i>Lucio Luzzatto, Lucia De Franceschi</i>	
101 Anemia Due to Acute Blood Loss	791
<i>Dan L. Longo</i>	
102 Bone Marrow Failure Syndromes Including Aplastic Anemia and Myelodysplasia.....	792
<i>Neal S. Young</i>	
103 Polycythemia Vera and Other Myeloproliferative Neoplasms	802
<i>Jerry L. Spivak</i>	
104 Acute Myeloid Leukemia	809
<i>William Blum</i>	
105 Chronic Myeloid Leukemia.....	818
<i>Hagop Kantarjian, Elias Jabbour, Jorge Cortes</i>	
106 Acute Lymphoid Leukemia	828
<i>Dieter Hoelzer</i>	
107 Chronic Lymphocytic Leukemia	834
<i>Jennifer A. Woyach, John C. Byrd</i>	
108 Non-Hodgkin's Lymphoma	841
<i>Caron A. Jacobson, Dan L. Longo</i>	
109 Hodgkin's Lymphoma.....	852
<i>Caron A. Jacobson, Dan L. Longo</i>	
110 Less Common Lymphoid and Myeloid Malignancies	855
<i>Ayalew Tefferi, Dan L. Longo</i>	
111 Plasma Cell Disorders.....	866
<i>Nikhil C. Munshi, Dan L. Longo, Kenneth C. Anderson</i>	
112 Amyloidosis	878
<i>John L. Berk, Vaishali Sanchorawala</i>	
113 Transfusion Therapy and Biology	884
<i>Pierre Tibergien, Olivier Garraud, Jacques Chiaroni</i>	
114 Hematopoietic Cell Transplantation.....	897
<i>Frederick R. Appelbaum</i>	

SECTION 3 Disorders of Hemostasis

115 Disorders of Platelets and Vessel Wall.....	903
<i>Barbara A. Konkle</i>	

116 Coagulation Disorders	910
<i>Jean M. Connors</i>	
117 Arterial and Venous Thrombosis	919
<i>Jane E. Freedman, Joseph Loscalzo</i>	
118 Antiplatelet, Anticoagulant, and Fibrinolytic Drugs	924
<i>Jeffrey I. Weitz</i>	

PART 5 Infectious Diseases

SECTION 1 Basic Considerations in Infectious Diseases

119 Approach to the Patient with an Infectious Disease.....	941
<i>Neeraj K. Surana, Dennis L. Kasper</i>	
120 Molecular Mechanisms of Microbial Pathogenesis.....	948
<i>Thomas E. Wood, Marcia B. Goldberg</i>	
121 Microbial Genomics and Infectious Disease	960
<i>Roby P. Bhattacharyya, Yonatan H. Grad, Deborah T. Hung</i>	
122 Approach to the Acutely Ill Infected Febrile Patient.....	973
<i>Tamar F. Barlam</i>	
123 Immunization Principles and Vaccine Use	981
<i>Sarah Mbaeyi, Amanda Cohn, Nancy Messonnier</i>	
124 Health Recommendations for International Travel	989
<i>Jesse Waggoner, Henry M. Wu</i>	
125 Climate Change and Infectious Disease.....	1001
<i>Aaron S. Bernstein</i>	

SECTION 2 Clinical Syndromes: Community-Acquired Infections

126 Pneumonia	1009
<i>Lionel A. Mandell, Michael S. Niederman</i>	
127 Lung Abscess.....	1020
<i>Rebecca M. Baron, Beverly W. Baron, Miriam Baron Barshak</i>	
128 Infective Endocarditis	1022
<i>Sara E. Cosgrove, Adolf W. Karchmer</i>	
129 Infections of the Skin, Muscles, and Soft Tissues	1034
<i>Dennis L. Stevens, Amy E. Bryant</i>	
130 Infectious Arthritis	1040
<i>Lawrence C. Madoff, Nongnooch Poowanawittayakom</i>	
131 Osteomyelitis.....	1046
<i>Werner Zimmerli</i>	
132 Intraabdominal Infections and Abscesses	1054
<i>Miriam Baron Barshak, Dennis L. Kasper</i>	
133 Acute Infectious Diarrheal Diseases and Bacterial Food Poisoning.....	1061
<i>Richelle C. Charles, Regina C. LaRocque</i>	
134 <i>Clostridioides difficile</i> Infection, Including Pseudomembranous Colitis.....	1066
<i>Dale N. Gerding, Stuart Johnson</i>	
135 Urinary Tract Infections, Pyelonephritis, and Prostatitis	1070
<i>Kalpana Gupta, Barbara W. Trautner</i>	
136 Sexually Transmitted Infections: Overview and Clinical Approach	1078
<i>Jeanne M. Marranzo, King K. Holmes</i>	
137 Encephalitis	1094
<i>Karen L. Roos, Michael R. Wilson, Kenneth L. Tyler</i>	
138 Acute Meningitis	1100
<i>Karen L. Roos, Kenneth L. Tyler</i>	

139 Chronic and Recurrent Meningitis	1110
<i>Avindra Nath, Walter J. Koroshetz, Michael R. Wilson</i>	
140 Brain Abscess and Empyema.....	1117
<i>Karen L. Roos, Kenneth L. Tyler</i>	
141 Infectious Complications of Bites.....	1124
<i>Sandeep S. Jubbal, Florencia Pereyra Segal, Lawrence C. Madoff</i>	
SECTION 3 Clinical Syndromes: Health Care-Associated Infections	
142 Infections Acquired in Health Care Facilities.....	1128
<i>Robert A. Weinstein</i>	
143 Infections in Transplant Recipients	1136
<i>Robert W. Finberg</i>	
SECTION 4 Therapy for Bacterial Diseases	
144 Treatment and Prophylaxis of Bacterial Infections	1148
<i>David C. Hooper, Erica S. Shenoy, Ramy H. Elshaboury</i>	
145 Bacterial Resistance to Antimicrobial Agents.....	1163
<i>David C. Hooper</i>	
SECTION 5 Diseases Caused by Gram-Positive Bacteria	
146 Pneumococcal Infections.....	1169
<i>David Goldblatt, Katherine L. O'Brien</i>	
147 Staphylococcal Infections.....	1178
<i>Franklin D. Lowy</i>	
148 Streptococcal Infections	1188
<i>Michael R. Wessels</i>	
149 Enterococcal Infections.....	1197
<i>William R. Miller, Cesar A. Arias, Barbara E. Murray</i>	
150 Diphtheria and Other Corynebacterial Infections.....	1203
<i>William R. Bishai, John R. Murphy</i>	
151 <i>Listeria monocytogenes</i> Infections.....	1208
<i>Jennifer P. Collins, Patricia M. Griffin</i>	
152 Tetanus.....	1211
<i>C. Louise Thwaites, Lam Minh Yen</i>	
153 Botulism.....	1214
<i>Carolina Lúquez, Jeremy Sobel</i>	
154 Gas Gangrene and Other Clostridial Infections.....	1220
<i>Amy E. Bryant, Dennis L. Stevens</i>	
SECTION 6 Diseases Caused by Gram-Negative Bacteria	
155 Meningococcal Infections	1225
<i>Manish Sadarangani, Andrew J. Pollard</i>	
156 Gonococcal Infections	1234
<i>Sanjay Ram, Peter A. Rice</i>	
157 <i>Haemophilus</i> and <i>Moraxella</i> Infections.....	1241
<i>Timothy F. Murphy</i>	
158 Infections Due to the HACEK Group and Miscellaneous Gram-Negative Bacteria.....	1246
<i>Tamar F. Barlam</i>	
159 <i>Legionella</i> Infections.....	1249
<i>Steven A. Pergam, Thomas R. Hawn</i>	
160 Pertussis and Other <i>Bordetella</i> Infections.....	1257
<i>Karina A. Top, Scott A. Halperin</i>	
161 Diseases Caused by Gram-Negative Enteric Bacilli	1261
<i>Thomas A. Russo, James R. Johnson</i>	
162 <i>Acinetobacter</i> Infections	1275
<i>Rossana Rosa, L. Silvia Munoz-Price</i>	
163 <i>Helicobacter pylori</i> Infections	1279
<i>John C. Atherton, Martin J. Blaser</i>	
164 Infections Due to <i>Pseudomonas</i> , <i>Burkholderia</i> , and <i>Stenotrophomonas</i> Species	1284
<i>Reuben Rampal</i>	
165 Salmonellosis	1291
<i>David A. Pegues, Samuel I. Miller</i>	
166 Shigellosis.....	1298
<i>Philippe J. Sansonetti, Jean Bergounioux</i>	
167 Infections Due to <i>Campylobacter</i> and Related Organisms	1302
<i>Martin J. Blaser</i>	
168 Cholera and Other Vibrioses.....	1305
<i>Matthew K. Waldor, Edward T. Ryan</i>	
169 Brucellosis	1310
<i>Nicholas J. Beeching</i>	
170 Tularemia	1314
<i>Max Maurin, Didier Raoult</i>	
171 Plague and Other <i>Yersinia</i> Infections	1320
<i>Michael B. Prentice</i>	
172 <i>Bartonella</i> Infections, Including Cat-Scratch Disease	1328
<i>Michael Giladi, Moshe Ephros</i>	
173 Donovanosis	1334
<i>Nigel O'Farrell</i>	
SECTION 7 Miscellaneous Bacterial Infections	
174 Nocardiosis.....	1335
<i>Gregory A. Filice</i>	
175 Actinomycosis.....	1340
<i>Thomas A. Russo</i>	
176 Whipple's Disease.....	1344
<i>Thomas A. Russo, Seth R. Glassman</i>	
177 Infections Due to Mixed Anaerobic Organisms	1347
<i>Neeraj K. Surana, Dennis L. Kasper</i>	
SECTION 8 Mycobacterial Diseases	
178 Tuberculosis.....	1357
<i>Mario C. Ravagliione, Andrea Gori</i>	
179 Leprosy.....	1382
<i>Jan H. Richardus, Hemanta K. Kar, Zoica Bakirtzief, Wim H. van Brakel</i>	
180 Nontuberculous Mycobacterial Infections	1392
<i>Steven M. Holland</i>	
181 Antimycobacterial Agents	1397
<i>Divya Reddy, Sebastian G. Kurz, Max R. O'Donnell</i>	
SECTION 9 Spirochetal Diseases	
182 Syphilis.....	1406
<i>Sheila A. Lukehart</i>	
183 Endemic Treponematoses	1413
<i>Sheila A. Lukehart, Lorenzo Giacani</i>	
184 Leptospirosis	1417
<i>Jiří F. P. Wagenaar, Marga G.A. Goris</i>	
185 Relapsing Fever and <i>Borrelia miyamotoi</i> Disease	1421
<i>Alan G. Barbour</i>	
186 Lyme Borreliosis	1425
<i>Allen C. Steere</i>	

SECTION 10 Diseases Caused by Rickettsiae, Mycoplasmas, and Chlamydiae	
187 Rickettsial Diseases.....	1431
<i>David H. Walker, J. Stephen Dumler, Lucas S. Blanton, Chantal P. Bleeker-Rovers</i>	
188 Infections Due to Mycoplasmas	1441
<i>R. Doug Hardy</i>	
189 Chlamydial Infections	1444
<i>Charlotte A. Gaydos, Thomas C. Quinn</i>	
SECTION 11 Viral Diseases: General Considerations	
190 Principles of Medical Virology	1453
<i>David M. Knipe</i>	
191 Antiviral Chemotherapy, Excluding Antiretroviral Drugs	1460
<i>Jeffrey I. Cohen, Eleanor Wilson</i>	
SECTION 12 Infections Due to DNA Viruses	
192 Herpes Simplex Virus Infections	1470
<i>Lawrence Corey</i>	
193 Varicella-Zoster Virus Infections.....	1479
<i>Richard J. Whitley</i>	
194 Epstein-Barr Virus Infections, Including Infectious Mononucleosis	1483
<i>Jeffrey I. Cohen</i>	
195 Cytomegalovirus and Human Herpesvirus Types 6, 7, and 8.....	1487
<i>Camille Nelson Kotton, Martin S. Hirsch</i>	
196 Molluscum Contagiosum, Monkeypox, and Other Poxvirus Infections	1492
<i>Inger K. Damon</i>	
197 Parvovirus Infections	1495
<i>Kevin E. Brown</i>	
198 Human Papillomavirus Infections	1498
<i>Darron R. Brown, Aaron C. Ermel</i>	
SECTION 13 Infections Due to DNA and RNA Respiratory Viruses	
199 Common Viral Respiratory Infections, Including COVID-19.....	1504
<i>James E. Crowe, Jr.</i>	
200 Influenza	1515
<i>Kathleen M. Neuzil, Peter F. Wright</i>	
SECTION 14 Infections Due to Human Immunodeficiency Virus and Other Human Retroviruses	
201 The Human Retroviruses	1521
<i>Dan L. Longo, Anthony S. Fauci</i>	
202 Human Immunodeficiency Virus Disease: AIDS and Related Disorders.....	1527
<i>Anthony S. Fauci, Gregory K. Folkers, H. Clifford Lane</i>	
SECTION 15 Infections Due to RNA Viruses	
203 Viral Gastroenteritis	1597
<i>Umesh D. Parashar, Roger I. Glass</i>	
204 Enterovirus, Parechovirus, and Reovirus Infections	1602
<i>Jeffrey I. Cohen</i>	
205 Measles (Rubeola).....	1608
<i>Kaitlin Rainwater-Lovett, William J. Moss</i>	
SECTION 16 Fungal Infections	
211 Pathogenesis, Diagnosis, and Treatment of Fungal Infections	1652
<i>Michail S. Lionakis, John E. Edwards Jr.</i>	
212 Histoplasmosis.....	1658
<i>Chadi A. Hage, L. Joseph Wheat</i>	
213 Coccidioidomycosis	1661
<i>Neil M. Ampel</i>	
214 Blastomycosis	1664
<i>Gregory M. Gauthier, Bruce S. Klein</i>	
215 Cryptococcosis.....	1668
<i>Arturo Casadevall</i>	
216 Candidiasis	1671
<i>Michail S. Lionakis, Shakti Singh, Ashraf S. Ibrahim, John E. Edwards, Jr.</i>	
217 Aspergillosis	1677
<i>David W. Denning</i>	
218 Mucormycosis.....	1681
<i>Brad Spellberg, Ashraf S. Ibrahim</i>	
219 Less Common Systemic Mycoses and Superficial Mycoses.....	1686
<i>Carol A. Kauffman</i>	
220 <i>Pneumocystis</i> Infections.....	1691
<i>Alison Morris, Henry Masur</i>	
SECTION 17 Protozoal and Helminthic Infections: General Considerations	
221 Introduction to Parasitic Infections	1696
<i>Sharon L. Reed</i>	
222 Agents Used to Treat Parasitic Infections.....	1701
<i>Thomas A. Moore</i>	
SECTION 18 Protozoal Infections	
223 Amebiasis and Infection with Free-Living Amebae.....	1714
<i>Rosa M. Andrade, Sharon L. Reed</i>	
224 Malaria.....	1720
<i>Nicholas J. White, Elizabeth A. Ashley</i>	
225 Babesiosis	1736
<i>Edouard Vannier, Jeffrey A. Gelfand</i>	
226 Leishmaniasis	1741
<i>Shyam Sundar</i>	
227 Chagas Disease and African Trypanosomiasis.....	1748
<i>François Chappuis, Yves Jackson</i>	
228 <i>Toxoplasma</i> Infections.....	1757
<i>Kami Kim</i>	
229 Protozoal Intestinal Infections and Trichomoniasis.....	1764
<i>Peter F. Weller</i>	

SECTION 19 Helminthic Infections

230 Introduction to Helminthic Infections.....	1768
Peter F. Weller	
231 Trichinellosis and Other Tissue Nematode Infections	1770
Peter F. Weller	
232 Intestinal Nematode Infections	1773
Thomas B. Nutman, Peter F. Weller	
233 Filarial and Related Infections.....	1778
Thomas B. Nutman, Peter F. Weller	
234 Schistosomiasis and Other Trematode Infections.....	1784
Birgitte Jyding Vennervold	
235 Cestode Infections	1790
A. Clinton White, Jr., Peter F. Weller	

PART 6 Disorders of the Cardiovascular System**SECTION 1 Introduction to Cardiovascular Disorders**

236 Approach to the Patient with Possible Cardiovascular Disease	1797
Joseph Loscalzo	
237 Basic Biology of the Cardiovascular System	1799
Joseph Loscalzo, John F. Keaney, Jr., Calum A. MacRae	
238 Epidemiology of Cardiovascular Disease	1810
Thomas A. Gaziano, J. Michael Gaziano	

SECTION 2 Diagnosis of Cardiovascular Disorders

239 Physical Examination of the Cardiovascular System.....	1815
Patrick T. O'Gara, Joseph Loscalzo	
240 Electrocardiography.....	1824
Ary L. Goldberger	
241 Noninvasive Cardiac Imaging: Echocardiography, Nuclear Cardiology, and Magnetic Resonance/ Computed Tomography Imaging.....	1832
Marcelo F. Di Carli, Raymond Y. Kwong, Scott D. Solomon	
242 Diagnostic Cardiac Catheterization and Coronary Angiography	1859
Jane A. Leopold, David P. Faxon	

SECTION 3 Disorders of Rhythm

243 Principles of Clinical Cardiac Electrophysiology	1866
William H. Sauer, Bruce A. Koplan, Paul C. Zei	
244 The Bradyarrhythmias: Disorders of the Sinoatrial Node.....	1873
William H. Sauer, Bruce A. Koplan	
245 The Bradyarrhythmias: Disorders of the Atrioventricular Node.....	1880
William H. Sauer, Bruce A. Koplan	
246 Approach to Supraventricular Tachyarrhythmias	1888
William H. Sauer, Paul C. Zei	
247 Physiologic and Nonphysiologic Sinus Tachycardia.....	1891
William H. Sauer, Paul C. Zei	
248 Focal Atrial Tachycardia.....	1893
William H. Sauer, Paul C. Zei	
249 Paroxysmal Supraventricular Tachycardias	1894
William H. Sauer, Paul C. Zei	

250 Common Atrial Flutter and Macroreentrant and Multifocal Atrial Tachycardias.....

1899
William H. Sauer, Paul C. Zei

251 Atrial Fibrillation.....

1903
William H. Sauer, Paul C. Zei

252 Approach to Ventricular Arrhythmias.....

1910
William H. Sauer, Usha B. Tedrow

253 Premature Ventricular Contractions, Nonsustained Ventricular Tachycardia, and Accelerated Idioventricular Rhythm.....

1915
William H. Sauer, Usha B. Tedrow

254 Sustained Ventricular Tachycardia

1919
William H. Sauer, Usha B. Tedrow

255 Polymorphic Ventricular Tachycardia and Ventricular Fibrillation

1923
William H. Sauer, Usha B. Tedrow

256 Electrical Storm and Incessant Ventricular Tachycardia.....

1927
William H. Sauer, Usha B. Tedrow

SECTION 4 Disorders of the Heart, Muscles, Valves, and Pericardium**257 Heart Failure: Pathophysiology and Diagnosis.....**

1930
Michael M. Givertz, Mandeep R. Mehra

258 Heart Failure: Management

1940
Akshay S. Desai, Mandeep R. Mehra

259 Cardiomyopathy and Myocarditis.....

1954
Neal K. Lakdawala, Lynne Warner Stevenson, Joseph Loscalzo

260 Cardiac Transplantation and Prolonged Assisted Circulation

1973
Mandeep R. Mehra

261 Aortic Stenosis.....

1978
Patrick T. O'Gara, Joseph Loscalzo

262 Aortic Regurgitation

1986
Patrick T. O'Gara, Joseph Loscalzo

263 Mitral Stenosis

1991
Patrick T. O'Gara, Joseph Loscalzo

264 Mitral Regurgitation

1995
Patrick T. O'Gara, Joseph Loscalzo

265 Mitral Valve Prolapse

1999
Patrick T. O'Gara, Joseph Loscalzo

266 Tricuspid Valve Disease

2001
Patrick T. O'Gara, Joseph Loscalzo

267 Pulmonic Valve Disease

2004
Patrick T. O'Gara, Joseph Loscalzo

268 Multiple and Mixed Valvular Heart Disease

2005
Patrick T. O'Gara, Joseph Loscalzo

269 Congenital Heart Disease in the Adult

2008
Anne Marie Valente, Michael J. Landzberg

270 Pericardial Disease

2019
Joseph Loscalzo

271 Atrial Myxoma and Other Cardiac Tumors

2025
Eric H. Awtry

272 Cardiac Trauma

2028
Eric H. Awtry

SECTION 5 Coronary and Peripheral Vascular Disease**273 Ischemic Heart Disease.....**

2030
Elliott M. Antman, Joseph Loscalzo

274 Non-ST-Segment Elevation Acute Coronary Syndrome (Non-ST-Segment Elevation Myocardial Infarction and Unstable Angina)	2046
<i>Robert P. Giugliano, Christopher P. Cannon, Eugene Braunwald</i>	
275 ST-Segment Elevation Myocardial Infarction	2053
<i>Elliott M. Antman, Joseph Loscalzo</i>	
276 Percutaneous Coronary Interventions and Other Interventional Procedures	2066
<i>David P. Faxon, Deepak L. Bhatt</i>	
277 Hypertension	2072
<i>Theodore A. Kotchen</i>	
278 Renovascular Disease	2088
<i>Stephen C. Textor</i>	
279 Deep-Venous Thrombosis and Pulmonary Thromboembolism	2091
<i>Samuel Z. Goldhaber</i>	
280 Diseases of the Aorta.....	2101
<i>Mark A. Creager, Joseph Loscalzo</i>	
281 Arterial Diseases of the Extremities.....	2107
<i>Mark A. Creager, Joseph Loscalzo</i>	
282 Chronic Venous Disease and Lymphedema	2115
<i>Mark A. Creager, Joseph Loscalzo</i>	
283 Pulmonary Hypertension	2121
<i>Bradley A. Maron, Joseph Loscalzo</i>	

PART 7 Disorders of the Respiratory System

SECTION 1 Diagnosis of Respiratory Disorders

284 Approach to the Patient with Disease of the Respiratory System	2131
<i>Bruce D. Levy</i>	
285 Disturbances of Respiratory Function	2133
<i>Edward T. Naureckas, Julian Solway</i>	
286 Diagnostic Procedures in Respiratory Disease	2140
<i>George R. Washko, Hilary J. Goldberg, Majid Shafiq</i>	

SECTION 2 Diseases of the Respiratory System

287 Asthma.....	2147
<i>Elliot Israel</i>	
288 Hypersensitivity Pneumonitis and Pulmonary Infiltrates with Eosinophilia.....	2160
<i>Praveen Akuthota, Michael E. Wechsler</i>	
289 Occupational and Environmental Lung Disease	2166
<i>John R. Balmes</i>	
290 Bronchiectasis.....	2173
<i>Rebecca M. Baron, Beverly W. Baron, Miriam Baron Barshak</i>	
291 Cystic Fibrosis	2176
<i>Eric J. Sorscher</i>	
292 Chronic Obstructive Pulmonary Disease.....	2180
<i>Edwin K. Silverman, James D. Crapo, Barry J. Make</i>	
293 Interstitial Lung Disease	2190
<i>Gary M. Hunninghake, Ivan O. Rosas</i>	
294 Disorders of the Pleura.....	2197
<i>Richard W. Light</i>	
295 Disorders of the Mediastinum	2200
<i>Richard W. Light</i>	
296 Disorders of Ventilation	2201
<i>John F. McConville, Julian Solway, Babak Mokhlesi</i>	

297 Sleep Apnea	2204
<i>Andrew Wellman, Daniel J. Gottlieb, Susan Redline</i>	
298 Lung Transplantation	2209
<i>Hilary J. Goldberg, Hari R. Mallidi</i>	
299 Interventional Pulmonary Medicine.....	2214
<i>Lonny Yarmus, David Feller-Kopman</i>	

PART 8 Critical Care Medicine

SECTION 1 Respiratory Critical Care

300 Approach to the Patient with Critical Illness.....	2217
<i>Rebecca M. Baron, Anthony F. Massaro</i>	
301 Acute Respiratory Distress Syndrome	2225
<i>Rebecca M. Baron, Bruce D. Levy</i>	
302 Mechanical Ventilatory Support.....	2230
<i>Scott Schissel</i>	

SECTION 2 Shock and Cardiac Arrest

303 Approach to the Patient with Shock	2235
<i>Anthony F. Massaro</i>	
304 Sepsis and Septic Shock	2241
<i>Emily B. Brant, Christopher W. Seymour, Derek C. Angus</i>	
305 Cardiogenic Shock and Pulmonary Edema	2250
<i>David H. Ingbar, Holger Thiele</i>	
306 Cardiovascular Collapse, Cardiac Arrest, and Sudden Cardiac Death	2257
<i>Christine Albert, William H. Sauer</i>	

SECTION 3 Neurologic Critical Care

307 Nervous System Disorders in Critical Care.....	2267
<i>J. Claude Hemphill, III, Wade S. Smith, S. Andrew Josephson, Daryl R. Gress</i>	

PART 9 Disorders of the Kidney and Urinary Tract

308 Approach to the Patient with Renal Disease or Urinary Tract Disease.....	2279
<i>Julian L. Seifter</i>	
309 Cell Biology and Physiology of the Kidney	2287
<i>Alfred L. George, Jr., Eric G. Neilson</i>	
310 Acute Kidney Injury.....	2296
<i>Sushrut S. Waikar, Joseph V. Bonventre</i>	
311 Chronic Kidney Disease	2309
<i>Joanne M. Bargman, Karl Skorecki</i>	
312 Dialysis in the Treatment of Kidney Failure	2320
<i>Kathleen D. Liu, Glenn M. Chertow</i>	
313 Transplantation in the Treatment of Renal Failure	2325
<i>Jamil Azzi, Naoka Murakami, Anil Chandraker</i>	
314 Glomerular Diseases	2331
<i>Julia B. Lewis, Eric G. Neilson</i>	
315 Polycystic Kidney Disease and Other Inherited Disorders of Tubule Growth and Development.....	2350
<i>Jing Zhou, Martin R. Pollak</i>	
316 Tubulointerstitial Diseases of the Kidney.....	2357
<i>Laurence H. Beck Jr., David J. Salant</i>	
317 Vascular Injury to the Kidney	2364
<i>Ronald S. Go, Nelson Leung</i>	

318 Nephrolithiasis	2368
<i>Gary C. Curhan</i>	
319 Urinary Tract Obstruction	2373
<i>Julian L. Seifter</i>	
320 Interventional Nephrology.....	2377
<i>Dirk M. Hentschel</i>	

PART 10 Disorders of the Gastrointestinal System

SECTION 1 Disorders of the Alimentary Tract

321 Approach to the Patient with Gastrointestinal Disease	2381
<i>William L. Hasler, Chung Owyang</i>	
322 Gastrointestinal Endoscopy	2387
<i>Louis Michel Wong Kee Song, Mark Topazian</i>	
323 Diseases of the Esophagus.....	2423
<i>Peter J. Kahrilas, Ikuo Hirano</i>	
324 Peptic Ulcer Disease and Related Disorders	2434
<i>John Del Valle</i>	
325 Disorders of Absorption.....	2458
<i>Deborah C. Rubin</i>	
326 Inflammatory Bowel Disease.....	2469
<i>Sonia Friedman, Richard S. Blumberg</i>	
327 Irritable Bowel Syndrome.....	2490
<i>Chung Owyang</i>	
328 Diverticular Disease and Common Anorectal Disorders.....	2497
<i>Susan L. Gearhart</i>	
329 Mesenteric Vascular Insufficiency.....	2506
<i>Maryam Ali Khan, Jaideep Das Gupta, Mahmoud Malas</i>	
330 Acute Intestinal Obstruction.....	2508
<i>Danny O. Jacobs</i>	
331 Acute Appendicitis and Peritonitis	2513
<i>Danny O. Jacobs</i>	

SECTION 2 Nutrition

332 Nutrient Requirements and Dietary Assessment.....	2517
<i>Johanna T. Dwyer</i>	
333 Vitamin and Trace Mineral Deficiency and Excess.....	2523
<i>Paolo M. Suter</i>	
334 Malnutrition and Nutritional Assessment	2534
<i>Gordon L. Jensen</i>	
335 Enteral and Parenteral Nutrition	2539
<i>L. John Hoffer, Bruce R. Bistrian, David F. Driscoll</i>	

SECTION 3 Liver and Biliary Tract Disease

336 Approach to the Patient with Liver Disease	2546
<i>Marc G. Ghany, Jay H. Hoofnagle</i>	
337 Evaluation of Liver Function.....	2553
<i>Emily D. Bethea, Daniel S. Pratt</i>	
338 The Hyperbilirubinemias	2557
<i>Allan W. Wolkoff</i>	
339 Acute Viral Hepatitis	2562
<i>Jules L. Dienstag</i>	
340 Toxic and Drug-Induced Hepatitis.....	2584
<i>William M. Lee, Jules L. Dienstag</i>	
341 Chronic Hepatitis	2591
<i>Jules L. Dienstag</i>	

342 Alcohol-Associated Liver Disease	2617
<i>Bernd Schnabl</i>	
343 Nonalcoholic Fatty Liver Diseases and Nonalcoholic Steatohepatitis	2619
<i>Manal F. Abdelmalek, Anna Mae Diehl</i>	
344 Cirrhosis and Its Complications	2624
<i>Alex S. Befeler, Bruce R. Bacon</i>	
345 Liver Transplantation.....	2633
<i>Raymond T. Chung, Jules L. Dienstag</i>	
346 Diseases of the Gallbladder and Bile Ducts.....	2641
<i>Norton J. Greenberger, Gustav Paumgartner, Daniel S. Pratt</i>	

SECTION 4 Disorders of the Pancreas

347 Approach to the Patient with Pancreatic Disease.....	2652
<i>Somashekhar G. Krishna, Darwin L. Conwell, Phil A. Hart</i>	
348 Acute and Chronic Pancreatitis	2657
<i>Phil A. Hart, Darwin L. Conwell, Somashekhar G. Krishna</i>	

PART 11 Immune-Mediated, Inflammatory, and Rheumatologic Disorders

SECTION 1 The Immune System in Health and Disease

349 Introduction to the Immune System	2671
<i>Barton F. Haynes, Kelly A. Soderberg, Anthony S. Fauci</i>	
350 Mechanisms of Regulation and Dysregulation of the Immune System.....	2701
<i>Barton F. Haynes, Kelly A. Soderberg, Anthony S. Fauci</i>	
351 Primary Immune Deficiency Diseases	2709
<i>Alain Fischer</i>	

SECTION 2 Disorders of Immune-Mediated Injury

352 Urticaria, Angioedema, and Allergic Rhinitis.....	2719
<i>Katherine L. Tuttle, Joshua A. Boyce</i>	
353 Anaphylaxis	2727
<i>David Hong, Joshua A. Boyce</i>	
354 Mastocytosis	2729
<i>Matthew P. Giannetti, Joshua A. Boyce</i>	
355 Autoimmunity and Autoimmune Diseases	2731
<i>Betty Diamond, Peter E. Lipsky</i>	
356 Systemic Lupus Erythematosus	2736
<i>Bevra Hannahs Hahn, Maureen McMahon</i>	
357 Antiphospholipid Syndrome	2749
<i>Haralampos M. Moutsopoulos, Clio P. Mavragani</i>	
358 Rheumatoid Arthritis.....	2751
<i>Ankoor Shah, E. William St. Clair</i>	
359 Acute Rheumatic Fever	2766
<i>Joseph Kado, Jonathan Carapetis</i>	
360 Systemic Sclerosis (Scleroderma) and Related Disorders	2771
<i>John Varga</i>	
361 Sjögren's Syndrome	2787
<i>Haralampos M. Moutsopoulos, Clio P. Mavragani</i>	
362 Spondyloarthritis	2790
<i>Joel D. Taurig, Lianne S. Gensler, Nigil Haroon</i>	
363 The Vasculitis Syndromes	2802
<i>Carol A. Langford, Anthony S. Fauci</i>	
364 Behçet Syndrome.....	2817

Yusuf Yazici

365 Inflammatory Myopathies.....	2819
<i>Steven A. Greenberg, Anthony A. Amato</i>	
366 Relapsing Polychondritis.....	2826
<i>Carol A. Langford</i>	
367 Sarcoidosis.....	2829
<i>Robert P. Baughman, Elyse E. Lower</i>	
368 IgG4-Related Disease.....	2837
<i>John H. Stone</i>	
369 Familial Mediterranean Fever and Other Hereditary Autoinflammatory Diseases.....	2840
<i>Daniel L. Kastner</i>	

SECTION 3 Disorders of the Joints and Adjacent Tissues

370 Approach to Articular and Musculoskeletal Disorders.....	2844
<i>John J. Cush</i>	
371 Osteoarthritis	2854
<i>David T. Felson, Tuhina Neogi</i>	
372 Gout and Other Crystal-Associated Arthropathies.....	2862
<i>Hyon K. Choi</i>	
373 Fibromyalgia.....	2868
<i>Leslie J. Crofford</i>	
374 Arthritis Associated with Systemic Disease, and Other Arthritides	2871
<i>Carol A. Langford, Brian F. Mandell</i>	
375 Periarticular Disorders of the Extremities	2878
<i>Carol A. Langford</i>	

PART 12 Endocrinology and Metabolism

SECTION 1 Endocrinology

376 Approach to the Patient with Endocrine Disorders	2881
<i>J. Larry Jameson</i>	
377 Mechanisms of Hormone Action	2884
<i>J. Larry Jameson</i>	
378 Physiology of Anterior Pituitary Hormones	2891
<i>Shlomo Melmed, J. Larry Jameson</i>	
379 Hypopituitarism	2896
<i>Shlomo Melmed, J. Larry Jameson</i>	
380 Pituitary Tumor Syndromes	2902
<i>Shlomo Melmed, J. Larry Jameson</i>	
381 Disorders of the Neurohypophysis	2918
<i>Gary L. Robertson, Daniel G. Bicket</i>	
382 Thyroid Gland Physiology and Testing.....	2926
<i>J. Larry Jameson, Susan J. Mandel, Anthony P. Weetman</i>	
383 Hypothyroidism.....	2933
<i>J. Larry Jameson, Susan J. Mandel, Anthony P. Weetman</i>	
384 Hyperthyroidism and Other Causes of Thyrotoxicosis	2938
<i>J. Larry Jameson, Susan J. Mandel, Anthony P. Weetman</i>	
385 Thyroid Nodular Disease and Thyroid Cancer.....	2946
<i>J. Larry Jameson, Susan J. Mandel, Anthony P. Weetman</i>	
386 Disorders of the Adrenal Cortex.....	2955
<i>Wiebke Arlt</i>	
387 Pheochromocytoma	2976
<i>Hartmut P. H. Neumann</i>	
388 Multiple Endocrine Neoplasia Syndromes.....	2983

R. V. Thakker

389 Autoimmune Polyendocrine Syndromes	2992
<i>Peter A. Gottlieb, Aaron W. Michels</i>	

SECTION 2 Sex- and Gender-Based Medicine

390 Sex Development	2997
<i>Courtney Finlayson, J. Larry Jameson, John C. Achermann</i>	
391 Disorders of the Testes and Male Reproductive System.....	3006
<i>Shalender Bhasin, J. Larry Jameson</i>	
392 Disorders of the Female Reproductive System	3027
<i>Janet E. Hall, Anuja Dokras</i>	
393 Menstrual Disorders and Pelvic Pain	3033
<i>Janet E. Hall, Anuja Dokras</i>	
394 Hirsutism	3039
<i>David A. Ehrmann</i>	
395 Menopause and Postmenopausal Hormone Therapy	3043
<i>JoAnn E. Manson, Shari S. Bassuk</i>	
396 Infertility and Contraception	3050
<i>Anuja Dokras, Janet E. Hall</i>	
397 Sexual Dysfunction	3055
<i>Kevin T. McVary</i>	
398 Women's Health.....	3063
<i>Emily Nosova, Andrea Dunaif</i>	
399 Men's Health	3069
<i>Shalender Bhasin</i>	
400 Lesbian, Gay, Bisexual, and Transgender (LGBT) Health	3078
<i>Baligh R. Yehia, Zachary B. R. McClain</i>	

SECTION 3 Obesity, Diabetes Mellitus, and Metabolic Syndrome

401 Pathobiology of Obesity.....	3080
<i>Stephen O'Rahilly, I. Sadaf Farooqi</i>	
402 Evaluation and Management of Obesity	3087
<i>Robert F. Kushner</i>	
403 Diabetes Mellitus: Diagnosis, Classification, and Pathophysiology	3094
<i>Alvin C. Powers, Kevin D. Niswender, Carmella Evans-Molina</i>	
404 Diabetes Mellitus: Management and Therapies.....	3104
<i>Alvin C. Powers, Michael J. Fowler, Michael R. Rickels</i>	
405 Diabetes Mellitus: Complications	3120
<i>Alvin C. Powers, John M. Stafford, Michael R. Rickels</i>	
406 Hypoglycemia	3129
<i>Stephen N. Davis, Philip E. Cryer</i>	
407 Disorders of Lipoprotein Metabolism.....	3135
<i>Daniel J. Rader</i>	
408 The Metabolic Syndrome.....	3150
<i>Robert H. Eckel</i>	

SECTION 4 Disorders of Bone and Mineral Metabolism

409 Bone and Mineral Metabolism in Health and Disease	3157
<i>F. Richard Bringhurst, Henry M. Kronenberg, Eva S. Liu</i>	
410 Disorders of the Parathyroid Gland and Calcium Homeostasis	3169
<i>John T. Potts, Jr., Harald Jüppner</i>	
411 Osteoporosis.....	3191

<i>Robert Lindsay, Blossom Samuels</i>	
412 Paget's Disease and Other Dysplasias of Bone	3209
<i>Rajesh K. Jain, Tamara J. Vokes</i>	
SECTION 5 Disorders of Intermediary Metabolism	
413 Heritable Disorders of Connective Tissue.....	3217
<i>Joan C. Marini, Fransiska Malfait</i>	
414 Hemochromatosis.....	3230
<i>Lawrie W. Powell, David M. Frazer</i>	
415 Wilson's Disease	3235
<i>Stephen G. Kaler</i>	
416 The Porphyrias	3237
<i>Robert J. Desnick, Manisha Balwani</i>	
417 Disorders of Purine and Pyrimidine Metabolism	3248
<i>John N. Mecchella, Christopher M. Burns</i>	
418 Lysosomal Storage Diseases	3254
<i>Robert J. Hopkin, Gregory A. Grabowski</i>	
419 Glycogen Storage Diseases and Other Inherited Disorders of Carbohydrate Metabolism.....	3261
<i>Priya S. Krishnani</i>	
420 Inherited Disorders of Amino Acid Metabolism in Adults.....	3268
<i>Nicola Longo</i>	
421 Inherited Defects of Membrane Transport	3274
<i>Nicola Longo</i>	

PART 13 Neurologic Disorders

SECTION 1 Diagnosis of Neurologic Disorders

422 Approach to the Patient with Neurologic Disease	3277
<i>Daniel H. Lowenstein, S. Andrew Josephson, Stephen L. Hauser</i>	
423 Neuroimaging in Neurologic Disorders	3282
<i>William P. Dillon</i>	
424 Pathobiology of Neurologic Diseases.....	3293
<i>Stephen L. Hauser, Arnold R. Kriegstein, Stanley B. Prusiner</i>	

SECTION 2 Diseases of the Central Nervous System

425 Seizures and Epilepsy.....	3305
<i>Vikram R. Rao, Daniel H. Lowenstein</i>	
426 Introduction to Cerebrovascular Diseases.....	3324
<i>Wade S. Smith, S. Claiborne Johnston, J. Claude Hemphill, III</i>	
427 Ischemic Stroke	3335
<i>Wade S. Smith, S. Claiborne Johnston, J. Claude Hemphill, III</i>	
428 Intracranial Hemorrhage.....	3348
<i>Wade S. Smith, J. Claude Hemphill, III, S. Claiborne Johnston</i>	
429 Subarachnoid Hemorrhage	3353
<i>J. Claude Hemphill, III, Wade S. Smith, Daryl R. Gress</i>	
430 Migraine and Other Primary Headache Disorders.....	3357
<i>Peter J. Goadsby</i>	
431 Alzheimer's Disease	3370
<i>Gil D. Rabinovici, William W. Seeley, Bruce L. Miller</i>	
432 Frontotemporal Dementia	3378
<i>William W. Seeley, Bruce L. Miller</i>	
433 Vascular Dementia	3381
<i>Steven M. Greenberg, William W. Seeley</i>	
434 Dementia with Lewy Bodies.....	3385

Irene Litvan, William W. Seeley, Bruce L. Miller

435 Parkinson's Disease	3386
<i>C. Warren Olanow, Anthony H.V. Schapira</i>	
436 Tremor, Chorea, and Other Movement Disorders	3400
<i>C. Warren Olanow, Christine Klein</i>	
437 Amyotrophic Lateral Sclerosis and Other Motor Neuron Diseases	3410
<i>Robert H. Brown, Jr.</i>	
438 Prion Diseases	3416
<i>Stanley B. Prusiner, Michael Geschwind</i>	
439 Ataxic Disorders	3422
<i>Roger N. Rosenberg</i>	
440 Disorders of the Autonomic Nervous System	3427
<i>Richard J. Barohn, John W. Engstrom</i>	
441 Trigeminal Neuralgia, Bell's Palsy, and Other Cranial Nerve Disorders	3436
<i>Vanja C. Douglas, Stephen L. Hauser</i>	
442 Diseases of the Spinal Cord	3445
<i>Stephen L. Hauser</i>	
443 Concussion and Other Traumatic Brain Injuries.....	3456
<i>Geoffrey T. Manley, Benjamin L. Brett, Michael McCrea</i>	
444 Multiple Sclerosis	3462
<i>Bruce A. C. Cree, Stephen L. Hauser</i>	
445 Neuromyelitis Optica	3477
<i>Bruce A. C. Cree, Stephen L. Hauser</i>	

SECTION 3 Nerve and Muscle Disorders

446 Peripheral Neuropathy	3480
<i>Anthony A. Amato, Richard J. Barohn</i>	
447 Guillain-Barré Syndrome and Other Immune-Mediated Neuropathies.....	3501
<i>Stephen L. Hauser, Anthony A. Amato</i>	
448 Myasthenia Gravis and Other Diseases of the Neuromuscular Junction	3509
<i>Anthony A. Amato</i>	
449 Muscular Dystrophies and Other Muscle Diseases	3516
<i>Anthony A. Amato, Robert H. Brown, Jr.</i>	

SECTION 4 Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

450 Myalgic Encephalomyelitis/Chronic Fatigue Syndrome	3532
<i>Elizabeth R. Unger, Jin-Mann S. Lin, Jeanne Bertolli</i>	

SECTION 5 Psychiatric and Addiction Disorders

451 Biology of Psychiatric Disorders	3534
<i>Robert O. Messing, Eric J. Nestler, Matthew W. State</i>	
452 Psychiatric Disorders	3540
<i>Victor I. Reus</i>	
453 Alcohol and Alcohol Use Disorders	3556
<i>Marc A. Schuckit</i>	
454 Nicotine Addiction	3563
<i>David M. Burns</i>	
455 Marijuana and Marijuana Use Disorders	3567
<i>Nora D. Volkow, Aidan Hampshire, Ruben Baler</i>	
456 Opioid-Related Disorders	3569
<i>Thomas R. Kosten, Colin N. Haile</i>	
457 Cocaine, Other Psychostimulants, and Hallucinogens.....	3573
<i>Karran A. Phillips, Wilson M. Compton</i>	

PART 14 Poisoning, Drug Overdose, and Envenomation

- 458 Heavy Metal Poisoning 3579
Howard Hu
- 459 Poisoning and Drug Overdose 3582
Mark B. Mycyk
- 460 Disorders Caused by Venomous Snakebites and Marine Animal Exposures 3596
Erik Fisher, Alex Chen, Charles Lei
- 461 Ectoparasite Infestations and Arthropod Injuries 3608
Richard J. Pollack, Scott A. Norton

PART 15 Disorders Associated with Environmental Exposures

- 462 Altitude Illness 3617
Buddha Basnyat, Geoffrey Tabin
- 463 Hyperbaric and Diving Medicine 3623
Michael H. Bennett, Simon J. Mitchell
- 464 Hypothermia and Peripheral Cold Injuries 3630
Daniel F. Danzl
- 465 Heat-Related Illnesses 3635
Daniel F. Danzl

PART 16 Genes, the Environment, and Disease

- 466 Principles of Human Genetics 3639
J. Larry Jameson, Peter Kopp
- 467 The Practice of Genetics in Clinical Medicine 3662
Susan M. Domchek, J. Larry Jameson, Susan Miesfeldt
- 468 Mitochondrial DNA and Heritable Traits and Diseases 3668
Karl L. Skorecki, Bruce H. Cohen
- 469 Telomere Disease 3680
Rodrigo T. Calado, Neal S. Young
- 470 Gene- and Cell-Based Therapy in Clinical Medicine 3685
Katherine A. High, Marcela V. Maus
- 471 The Human Microbiome 3690
Neeraj K. Surana, Dennis L. Kasper

PART 17 Global Medicine

- 472 Global Issues in Medicine 3703
Joseph J. Rhatigan, Paul Farmer
- 473 Emerging and Reemerging Infectious Diseases 3713
George W. Rutherford, Jaime Sepúlveda
- 474 Primary Care and Global Health 3718
Tim Evans, Kumaran Rasanathan
- 475 Health Effects of Climate Change 3726
Eugene T. Richardson, Maxine A. Burkett, Paul E. Farmer

PART 18 Aging

- 476 Biology of Aging 3733
Rafael de Cabo, David Le Couteur
- 477 Caring for the Geriatric Patient 3739
Joseph G. Ouslander, Bernardo Reyes

PART 19 Consultative Medicine

- 478 Approach to Medical Consultation 3761
Jeffrey Berns, Jack Ende
- 479 Medical Disorders During Pregnancy 3762
Sarah Rae Easter, Robert L. Barbieri
- 480 Medical Evaluation of the Surgical Patient 3769
Prashant Vaishnava, Kim A. Eagle

PART 20 Frontiers

- 481 Behavioral Economics and Health 3775
Kevin G. Volpp, George Loewenstein, David A. Asch
- 482 Complementary and Integrative Therapies and Practices 3784
Helene M. Langevin
- 483 The Role of Epigenetics in Disease and Treatment 3790
Brian C. Capell, Shelley L. Berger
- 484 Applications of Stem Cell Biology in Clinical Medicine 3796
John A. Kessler
- 485 The Role of Circadian Biology in Health and Disease 3800
Jonathan Cedernaes, Kathryn Moynihan Ramsey, Joseph Bass
- 486 Network Medicine: Systems Biology in Health and Disease 3812
Joseph Loscalzo
- 487 Emerging Neurotherapeutic Technologies 3819
Jyoti Mishra, Karunesh Ganguly
- 488 Machine Learning and Augmented Intelligence in Clinical Medicine 3826
Arjun K. Manrai, Isaac S. Kohane
- 489 Metabolomics 3831
Jared R. Mayers, Mathew G. Vander Heiden
- 490 Circulating Nucleic Acids as Liquid Biopsies and Noninvasive Disease Biomarkers 3836
Ash A. Alizadeh, Kiran K. Khush, Yair J. Blumenfeld
- 491 Protein Folding Disorders 3846
Richard I. Morimoto, G. Scott Budinger
- 492 Novel Approaches to Diseases of Unknown Etiology 3851
David Adams, Camilo Toro, Joseph Loscalzo
- Index I-1

Related Harrison's 21st Edition Content

The following chapters are available online. They can be viewed by opening the table of contents of *Harrison's 21st edition* at www.accessmedicine.com/harrisons.

Video Collection

V1 Video Library of Gait Disorders

Gail Kang, Nicholas B. Galifianakis, Michael D. Geschwind

V2 Primary Progressive Aphasia, Memory Loss, and Other Focal Cerebral Disorders

Maria Luisa Gorno-Tempini, Jennifer Ogar, Joel Kramer, Bruce L. Miller, Gil D. Rabinovici, Maria Carmela Tartaglia

V3 Video Library of Neuro-Ophthalmology

Jonathan C. Horton

V4 Examination of the Comatose Patient

S. Andrew Josephson

V5 Video Atlas of Gastrointestinal Endoscopic Lesions

Louis Michel Wong Kee Song, Mark Topazian

V6 The Neurologic Screening Exam

Daniel H. Lowenstein

V7 Video Atlas of the Detailed Neurologic Examination

Martin A. Samuels

Supplementary Topics

S1 Fluid and Electrolyte Imbalances and Acid-Base Disturbances: Case Examples

David B. Mount, Thomas D. DuBose, Jr.

S2 Cerebrospinal Fluid Disturbances: Case Examples

Prashanth S. Ramachandran, Michael R. Wilson

S3 Microbial Bioterrorism

H. Clifford Lane, Anthony S. Fauci

S4 Chemical Terrorism

James A. Romano, Jr., Jonathan Newmark

S5 Radiation Terrorism

Christine E. Hill-Kayser, Eli Glatstein, Zelig A. Tochner

S6 Infections in War Veterans

Andrew W. Artenstein

S7 Health Care for Military Veterans

Stephen C. Hunt, Lucile Burgo-Black, Charles W. Hoge

S8 Primary Immunodeficiencies Associated with (or Secondary to) Other Diseases

Alain Fischer

S9 Technique of Lumbar Puncture

Elizabeth Robbins, Stephen L. Hauser

S10 The Clinical Laboratory in Modern Health Care

Anthony A. Killeen

S11 Laboratory Diagnosis of Infectious Diseases

Manfred Brigl, Alexander J. McAdam

S12 Laboratory Diagnosis of Parasitic Infections

Sharon L. Reed, Sanjay R. Mehta

Atlases

A1 Atlas of Rashes Associated with Fever

Kenneth M. Kaye, Elaine T. Kaye

A2 Atlas of Blood Smears of Malaria and Babesiosis

Nicholas J. White, Elizabeth A. Ashley

A3 Atlas of Oral Manifestations of Disease

Samuel C. Durso, Janet A. Yellowitz

A4 Atlas of Urinary Sediments and Renal Biopsies

Agnes B. Fogo, Eric G. Neilson

A5 Atlas of Skin Manifestations of Internal Disease

Thomas J. Lawley, Benjamin K. Stoff, Calvin O. McCall

A6 Atlas of Hematology

Dan L. Longo

A7 Atlas of Electrocardiography

Ary L. Goldberger

A8 Atlas of Cardiac Arrhythmias

Ary L. Goldberger

A9 Atlas of Noninvasive Imaging

Marcelo F. Di Carli, Raymond Y. Kwong, Scott D. Solomon

A10 Atlas of Atherosclerosis

Peter Libby

A11 Atlas of Percutaneous Revascularization and Adult Structural Heart Interventions

Jane A. Leopold, Deepak L. Bhatt, David P. Faxon

A12 Atlas of Chest Imaging

Samuel Y. Ash, George R. Washko

A13 Atlas of Liver Biopsies

Jules L. Dienstag, Atul K. Bhan

A14 Atlas of the Vasculitic Syndromes

Carol A. Langford, Anthony S. Fauci

A15 Atlas of Clinical Manifestations of Endocrine and Metabolic Diseases

J. Larry Jameson

A16 Atlas of Neuroimaging

Michael F. Regner, Andre D. Furtado, Luciano Villarinho, William P. Dillon

Contributors

James L. Abbruzzese, MD, FACP, FASCO, DSc (hon)

Professor, Division of Medical Oncology, Duke Cancer Institute, Durham, North Carolina [92]

Manal F. Abdelmalek, MD, MPH

Professor of Medicine, Division of Gastroenterology and Hepatology, Duke University, Durham, North Carolina [343]

John C. Achermann, MD, PhD

Wellcome Trust Senior Research Fellow in Clinical Science, Genetics & Genomic Medicine, UCL GOS Institute of Child Health, University College London, London, United Kingdom [390]

David Adams, MD, PhD

Deputy Director of Clinical Genomics, Office of the Clinical Director/NHGRI and Undiagnosed Diseases Program, National Institutes of Health, Bethesda, Maryland [492]

John W. Adamson, MD

Clinical Professor, Division of Hematology/Oncology, Department of Medicine, University of California at San Diego, San Diego, California [63, 97]

Praveen Akuthota, MD

Associate Clinical Professor, Division of Pulmonary, Critical Care & Sleep Medicine, University of California, San Diego, San Diego, California [288]

Christine Albert, MD, MPH

Chair, Department of Cardiology; Lee and Harold Kapelovitz Endowed Chair in Research Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California [306]

Ash A. Alizadeh, MD, PhD

Professor of Medicine (Oncology), Stanford University School of Medicine, Stanford, California [490]

Anthony A. Amato, MD

Professor of Neurology, Harvard Medical School; Distinguished Chair of Neurology and Chief, Neuromuscular Division, Brigham and Women's Hospital, Boston, Massachusetts [365, 446–449]

Rachel L. Amdur, MD

Assistant Professor of Medicine, Division of General Internal Medicine and Geriatrics, Northwestern University Feinberg School of Medicine, Chicago, Illinois [35]

Neil M. Ampel, MD

Professor Emeritus of Medicine and Immunobiology, University of Arizona, Tucson, Arizona [213]

Kenneth C. Anderson, MD

Kraft Family Professor of Medicine, Harvard Medical School; Chief, Jerome Lipper Multiple Myeloma Center, Dana-Farber Cancer Institute, Boston, Massachusetts [111]

Rosa M. Andrade, MD

Assistant Professor of Medicine, University of California, Irvine School of Medicine, Irvine, California [223]

Derek C. Angus, MD, MPH

Distinguished Professor and Mitchell P. Fink Endowed Chair, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania [304]

Elliott M. Antman, MD

Professor of Medicine; Harvard Medical School; Senior Physician; Senior Investigator, TIMI Study Trial, Brigham and Women's Hospital, Boston, Massachusetts [273, 275]

Frederick R. Appelbaum, MD

Deputy Director, Fred Hutchinson Cancer Research Center, Seattle, Washington [114]

Cesar A. Arias, MD, PhD, MSc, FIDSA

Chief, Division of Infectious Diseases, Houston Methodist Hospital; Professor and John F. III and Ann H. Bookout Distinguished Chair; Co-Director Center for Infectious Diseases Research, Houston Methodist Research Institute and Weill Cornell Medical College, Houston, Texas [149]

Wiebke Arlt, MD, DSc, FRCP, FMedSci

William Withering Chair of Medicine, Institute of Metabolism and Systems Research, University of Birmingham; Consultant Endocrinologist, Queen Elizabeth Hospital Birmingham, Birmingham, United Kingdom [386]

Katrina A. Armstrong, MD

Physician in Chief, Massachusetts General Hospital, Boston, Massachusetts [6]

Andrew W. Artenstein, MD

Chief Physician Executive and Chief Academic Officer, Baystate Health; Regional Executive Dean and Professor of Medicine, University of Massachusetts Chan Medical School-Baystate, Springfield, Massachusetts [S6]

David A. Asch, MD, MBA

Executive Director, Penn Medicine Center for Health Care Innovation; John Morgan Professor, Perelman School of Medicine and the Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania [481]

Samuel Y. Ash, MD, MPH

Assistant Professor of Medicine, Harvard Medical School; Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [A12]

Elizabeth A. Ashley, MB, BS, MRCP, FRCPath

Professor of Tropical Medicine, Oxford University; Director, Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit, Vientiane, Lao PDR [224, A2]

John C. Atherton, MD, FRCP

Professor of Gastroenterology and Dean of the Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, United Kingdom [163]

Eric H. Awtry, MD

Associate Professor of Medicine, Boston University School of Medicine; Associate Chair for Clinical Affairs, Section of Cardiology, Boston Medical Center, Boston, Massachusetts [271, 272]

Jamil Azzi, MD

Associate Physician, Renal Division, Brigham and Women's Hospital; Director, Renal Transplant Fellowship; Assistant Professor of Medicine, Harvard Medical School, Boston, Massachusetts [313]

Bruce R. Bacon, MD

Emeritus Professor of Internal Medicine, Saint Louis University School of Medicine, St. Louis, Missouri [344]

Jessica M. Baker, MD

Assistant Professor, Department of Neurology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin [26]

Zoica Bakirtzief, PhD

Guest Professor of Psychology, Vocational Teacher Certification Program, Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brazil [179]

Ruben Baler, PhD

Health Scientist, Office of Science Policy and Communications, National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland [455]

John R. Balmes, MD

Professor of Medicine, University of California San Francisco School of Medicine, San Francisco, California [289]

Manisha Balwani, MD, MS

Professor, Department of Genetics and Genomic Sciences and Medicine, Icahn School of Medicine at Mount Sinai, New York, New York [416]

Robert L. Barbieri, MD

Kate Macy Ladd Distinguished Professor of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School; Chief of Obstetrics, Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, Massachusetts [479]

Alan G. Barbour, MD

Distinguished Professor of Medicine and Microbiology and Molecular Genetics, University of California, Irvine, Irvine, California [185]

Joanne M. Bargman, MD, FRCPC

Professor of Medicine, University of Toronto; Staff Nephrologist, University Health Network; Clinician Investigator, Toronto General Hospital Research Institute; Director, Peritoneal Dialysis Program, Co-Director, Renal-Rheumatology Lupus Clinic, University Health Network [311]

Tamar F. Barlam, MD, MSc

Professor of Medicine, Boston University School of Medicine; Chief, Section of Infectious Diseases, Boston Medical Center, Boston, Massachusetts [122, 158]

Richard J. Barohn, MD

Executive Vice Chancellor for Health Affairs; Executive Director, NextGen Precision Health, University of Missouri, Columbia, Missouri [440, 446]

Beverly W. Baron, MD

Professor of Pathology, Retired, University of Chicago, Chicago, Illinois [127, 290]

Rebecca M. Baron, MD

Associate Professor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Pulmonary Division and Critical Care, Boston, Massachusetts [37, 127, 290, 300, 301]

Miriam Baron Barshak, MD

Assistant Professor of Medicine, Harvard Medical School; Physician, Massachusetts General Hospital, Boston, Massachusetts [127, 132, 290]

Buddha Basnyat, MSc, MD, FACP, FRCP(Edinburgh)

Director, Oxford University Clinical Research Unit—Nepal, Patan Hospital, Kathmandu, Nepal [462]

Joseph Bass, MD, PhD

Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Feinberg School of Medicine, Department of Neurobiology, Northwestern University, Chicago, Illinois [485]

Shari S. Bassuk, ScD

Epidemiologist, Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts [395]

David W. Bates, MD, MSc

Professor of Medicine, Harvard Medical School; Chief, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Phyllis Jen Center for Primary Care, Boston, Massachusetts [8]

Robert P. Baughman, MD

Department of Internal Medicine, University of Cincinnati Medical Center, Cincinnati, Ohio [367]

Laurence H. Beck, Jr., MD, PhD

Associate Professor of Medicine, Boston University School of Medicine, Boston, Massachusetts [316]

Nicholas J. Beeching, FRCP, FRACP, FFTM RCPS(Glasg), FESCMID, FISTM, DTM&H, DCH

Consultant in Tropical and Infectious Diseases, Tropical and Infectious Disease Unit, Royal Liverpool University Hospitals Foundation NHS Trust; Emeritus Professor of Tropical and Infectious Diseases, Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom [169]

Alex S. Befeler, MD

Medical Director Liver Transplant, Professor of Internal Medicine, Division of Gastroenterology and Hepatology, Saint Louis University, St. Louis, Missouri [344]

Michael H. Bennett, MD, MBBS, MM (Clin Epi)

Conjoint Professor in Anesthesia and Hyperbaric Medicine; Faculty of Medicine, University of New South Wales; Academic Head of Department, Wales Anaesthesia, Prince of Wales Hospital, Sydney, Australia [463]

Shelley L. Berger, PhD

Daniel S. Och University Professor, Departments of Cell and Developmental Biology; Biology; Genetics; Director, Penn Epigenetics Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania [483]

Jean Bergounioux, MD, PhD

Professor of Medicine, Versailles Saint Quentin University - Paris Saclay, UFR Simone Veil - Motigny le Bretonneux, France; Director, Department of Pediatric Neurology and Intensive Care Medicine, Assistance Publique des Hôpitaux de Paris, Garches, France [166]

John L. Berk, MD

Professor of Medicine, Boston University School of Medicine, Assistant Director, Amyloidosis Center, Boston Medical Center, Boston, Massachusetts [112]

Jeffrey Berns, MD

Professor of Medicine and Pediatrics; Associate Chief, Renal Electrolyte and Hypertension Division; Vice-President and Associate Dean for Graduate Medical Education, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, Pennsylvania [478]

Aaron S. Bernstein, MD, MPH

Assistant Professor of Pediatrics, Harvard Medical School; Hospitalist, Division of General Pediatrics, Boston Children's Hospital; Interim Director, Center for Climate, Health and the Global Environment, Harvard T.H. Chan School of Public Health, Boston, Massachusetts [125]

Jeanne Bertolli, PhD

Division of High-Consequence Pathogens and Pathology, National Center for Zoonotic and Emerging Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [450]

Joseph R. Betancourt, MD, MPH

Associate Professor of Medicine, Massachusetts General Hospital; Harvard Medical School, Boston, Massachusetts [10]

Emily D. Bethea, MD

Instructor in Medicine, Harvard Medical School; Associate Clinical Director of Liver Transplantation, Gastroenterology and Hepatology Division, Massachusetts General Hospital, Boston, Massachusetts [337]

Julie A. Bettinger, MPH, PhD

Professor, Department of Pediatrics, Vaccine Evaluation Center, BC Children's Hospital, University of British Columbia, Vancouver, British Columbia, Canada [3]

Atul K. Bhan, MBBS, MD

Professor of Pathology, Harvard Medical School, Associate Director, Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital, Boston, Massachusetts [A13]

Shalender Bhasin, MB, BS

Professor of Medicine, Harvard Medical School; Director, Research Program in Men's Health: Aging and Metabolism; Director, Boston Claude D. Pepper Older Americans Independence Center; Brigham and Women's Hospital, Boston, Massachusetts [391, 399]

Deepak L. Bhatt, MD, MPH, FACC, FAHA, FSCAI, FESC

Professor of Medicine, Harvard Medical School; Executive Director of Interventional Cardiovascular Programs, Brigham and Women's Hospital Heart & Vascular Center, Boston, Massachusetts [276, A11]

Roby P. Bhattacharyya, MD, PhD

Assistant Professor of Medicine, Harvard Medical School and Massachusetts General Hospital; Associate Member, Broad Institute of MIT and Harvard, Boston, Massachusetts [121]

Daniel G. Bichel, MD

Professor of Medicine, Pharmacology and Physiology, University of Montreal; Staff Nephrologist, Hôpital du Sacré-Cœur de Montréal, Montréal, Quebec, Canada [381]

David R. Bickers, MD

Carl Truman Nelson Professor and Chair, Department of Dermatology, Columbia University Irving Medical Center, New York, New York [61]

William R. Bishai, MD, PhD

Professor of Medicine, Division of Infectious Diseases, Johns Hopkins School of Medicine, Baltimore, Maryland [150]

Bruce R. Bistrian, MD, PhD, MPH

Professor of Medicine, Harvard Medical School; Chief, Clinical Nutrition, Beth Israel Deaconess Medical Center, Boston, Massachusetts [335]

Lucas S. Blanton, MD

Assistant Professor, Division of Infectious Diseases, Department of Internal Medicine University of Texas Medical Branch, Galveston, Texas [187]

Martin J. Blaser, MD

Henry Rutgers Chair of the Human Microbiome; Director, Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, New Jersey [163, 167]

Chantal P. Bleeker-Rovers, MD, PhD

Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands [20, 187]

William Blum, MD

Director, Acute Leukemia Program; Professor, Department of Hematology and Oncology, Winship Cancer Institute and Emory University, Atlanta, Georgia [104]

Richard S. Blumberg, MD

Vice-Chair for Research in Department of Medicine, Brigham and Women's Hospital, Professor of Medicine, Harvard Medical School, Boston, Massachusetts [326]

Yair J. Blumenfeld, MD

Associate Professor of Obstetrics and Gynecology (Maternal Fetal Medicine), Stanford University School of Medicine, Stanford, California [490]

Jean L. Bolognia, MD

Professor, Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut [58]

Joseph V. Bonventre, MD, PhD

Chief, Renal Division and Engineering in Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [310]

Joshua A. Boyce, MD

Professor of Medicine and Pediatrics; Albert L. Sheffer Professor of Medicine, Harvard Medical School; Director, Inflammation and Allergic Disease Research Section, Brigham and Women's Hospital, Boston, Massachusetts [352–354]

Emily B. Brant, MD, MS

Assistant Professor, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania [304]

Eugene Braunwald, MD

Distinguished Hersey Professor of the Theory and Practice of Medicine, Harvard Medical School; Brigham and Women's Hospital, BWH/Founding Chair, TIMI Group, Boston, Massachusetts [274]

Irwin M. Braverman, MD

Professor Emeritus; Senior Research Scientist, Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut [58]

Otis W. Brawley, MD, MACP, FRCP(L), FASCO, FACE

Bloomberg Distinguished Professor, Johns Hopkins School of Medicine and Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland [70]

Benjamin L. Brett, PhD

Medical College of Wisconsin, Assistant Professor, Departments of Neurosurgery and Neurology (Division Neuropsychology), Milwaukee, Wisconsin [443]

Manfred Brigl, MD

Assistant Professor of Pathology, Harvard Medical School, Boston, Massachusetts [S11]

F. Richard Bringhurst, MD

Associate Professor of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts [409]

Steven M. Bromley, MD

Director, South Jersey MS Center, Bromley Neurology PC, Audubon, New Jersey [33]

Darron R. Brown, MD

Professor of Medicine, Microbiology, and Immunology, Division of Infectious Diseases, Indiana University School of Medicine, Indianapolis, Indiana [198]

Kevin E. Brown, MD, MRCP, FRCPath

Consultant Medical Virologist, Immunisation and Vaccine Preventable Diseases Division, UK Health Security Agency, London, United Kingdom [197]

Robert H. Brown, Jr., MD, PhD

Chairman, Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts [437, 449]

Amy E. Bryant, PhD

Research Professor, Department of Biomedical and Pharmaceutical Sciences College of Pharmacy, Idaho State University, Meridian, Idaho [129, 154]

G. Scott Budinger, MD

Ernest S. Bazley Professor of Airway Diseases; Chief, Pulmonary and Critical Care Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [491]

Fred Bunz, MD, PhD

Associate Professor, Johns Hopkins University School of Medicine, Baltimore, Maryland [71]

Lucile Burgo-Black, MD, FACP

National Co-Director, VA Post-Deployment Integrated Care Initiative, Assistant Clinical Professor of Medicine, Department of General Internal Medicine, Yale University School of Medicine, New Haven, Connecticut [S7]

Maxine A. Burkett, JD

Professor of Law, William S. Richardson School of Law, University of Hawaii at Mānoa, Honolulu, Hawaii [475]

Christopher M. Burns, MD

Associate Professor of Medicine, Geisel School of Medicine at Dartmouth, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire [417]

David M. Burns, MD

Professor Emeritus, University of California, San Diego School of Medicine Del Mar, California [454]

John C. Byrd, MD

D. Warren Brown Chair of Leukemia Research; Distinguished University Professor of Medicine, Medicinal Chemistry, and Veterinary Biosciences; Director, Division of Hematology, Department of Medicine, The Ohio State University, Columbus, Ohio [107]

Rodrigo T. Calado, MD, PhD

Associate Professor of Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil [469]

Michael Camilleri, MD

Atherton and Winifred W. Bean Professor; Professor of Medicine, Pharmacology, and Physiology, Mayo Clinic School of Medicine, Rochester, Minnesota [46]

Christopher P. Cannon, MD

Professor of Medicine, Harvard Medical School; Education Director, Cardiovascular Innovation, Preventive Cardiology Section, Brigham and Women's Hospital, Boston, Massachusetts [274]

Brian C. Capell, MD, PhD

Assistant Professor of Dermatology and Genetics, Departments of Dermatology and Genetics, Penn Epigenetics Institute, Abramson Cancer Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania [483]

Jonathan Carapetis, MBBS, FRACP, FAFPHM, PhD

Executive Director, Telethon Kids Institute, Perth Children's Hospital, Nedlands, Western Australia [359]

Arturo Casadevall, MD, PhD

Professor and Chair, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland [215]

Jonathan Cedernaes, MD, PhD

Visiting Postdoctoral Fellow, Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [485]

Anil K. Chadraker, MB, ChB

Associate Professor of Medicine, Harvard Medical School; Medical Director, Kidney and Pancreas Transplantation, Brigham and Women's Hospital; Boston, Massachusetts [313]

François Chappuis, MD, MCTM, PhD

Head of Division, Division of Tropical and Humanitarian Medicine, Geneva University Hospitals, Geneva, Switzerland [227]

Richelle C. Charles, MD, FIDSA

Associate Professor of Medicine, Harvard Medical School, Massachusetts General Hospital; Associate Professor of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts [133]

Justin T. Cheeley, MD, FAAD

Assistant Professor, Divisions of Dermatology and Internal Medicine and Geriatrics, Emory University School of Medicine, Atlanta, Georgia [57]

Alex Chen, MD

Associate Physician, Department of Emergency Medicine, Kaiser Permanente, South Sacramento Campus, Sacramento, California [460]

Glenn M. Chertow, MD, MPH

Professor of Medicine, Division of Nephrology, Stanford University School of Medicine, Palo Alto, California [312]

Jacques Chiaroni, MD, PhD

Professor, Aix Marseille Univ, CNRS, EFS, ADES, UMR 7268; Etablissement Français du Sang Provence Alpes Côte d'Azur et Corse, Marseille, France [113]

Hyon K. Choi, MD, DrPH

Professor of Medicine, Harvard Medical School; Director, Gout and Crystal Arthropathy Center; Director, Clinical Epidemiology and Health Outcomes, Division of Rheumatology, Allergy, and Immunology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts [372]

Benjamin F. Chong, MD, MSCS

Associate Professor, Department of Dermatology, University of Texas Southwestern Medical Center, Dallas, Texas [59]

Raymond T. Chung, MD

Professor of Medicine, Harvard Medical School; Director of Hepatology and Liver Center; Vice Chief, Gastroenterology Division; Kevin and Polly Maroni MGH Research Scholar, Massachusetts General Hospital, Boston, Massachusetts [345]

Jeffrey W. Clark, MD

Associate Professor of Medicine, Harvard Medical School; Medical Director, Clinical Trials Core, Dana-Farber Harvard Cancer Center; Massachusetts General Hospital, Boston, Massachusetts [72]

Bruce H. Cohen, MD, FAAN

Professor of Pediatrics, Northeast Ohio Medical University; Professor of Integrative Medical Sciences, Northeast Ohio Medical University; Considine Family Endowed Chair in Research – Akron Children's Hospital; Director, NeuroDevelopmental Science Center, Akron Children's Hospital; Divisions of Neurology, Neurosurgery, NeuroBehavioral Health, Psychiatry and Developmental Pediatrics; Interim Vice President and Medical Director, Rebecca D. Considine Research Institute, Akron Children's Hospital, Akron, Ohio [468]

Jeffrey I. Cohen, MD

Chief, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [191, 194, 204]

Amanda Cohn, MD

Chief Medical Officer, National Center for Immunization and Respiratory Diseases (NCIRD), Atlanta, Georgia [123]

Jennifer P. Collins, MD, MSc

Enteric Diseases Epidemiology Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [151]

Wilson M. Compton, MD, MPE

Deputy Director, National Institute on Drug Abuse, National Institutes of Health, U.S. Department of Health and Human Services, Bethesda, Maryland [457]

Jean M. Connors, MD

Hematology Division, Brigham and Women's Hospital; Harvard Medical School, Boston, Massachusetts [116]

Darwin L. Conwell, MD, MS

Professor of Medicine, The Ohio State University College of Medicine; Director, Division of Gastroenterology, Hepatology and Nutrition; The Ohio State University Wexner Medical Center, Columbus, Ohio [347, 348]

Lawrence Corey, MD

Professor of Medicine and Laboratory Medicine and Pathology, University of Washington; Past President & Director, Fred Hutchinson Cancer Research Center; Professor, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington [192]

Jorge Cortes, MD

Jane and John Justin Distinguished Chair in Leukemia Research; Deputy Chairman; Section Chief of AML and CML, The University of Texas MD Anderson Cancer Center, Houston, Texas [105]

Sara E. Cosgrove, MD, MS

Professor of Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, Maryland [128]

James D. Crapo, MD

Professor of Medicine, Department of Medicine; Division of Pulmonary and Critical Care & Sleep Medicine, National Jewish Health, Denver, Colorado [292]

Mark A. Creager, MD

Professor of Medicine, Professor of Surgery, Geisel School of Medicine at Dartmouth; Director, Heart and Vascular Center, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire [280–282]

Bruce A. C. Cree, MD, PhD, MAS

George A. Zimmermann Endowed Professor in Multiple Sclerosis; Professor of Clinical Neurology; Clinical Research Director, UCSF Weill Institute for Neurosciences, Department of Neurology, University of California San Francisco, San Francisco, California [444, 445]

Leslie J. Crofford, MD

Professor, Departments of Medicine and Pathology, Microbiology and Immunology, Vanderbilt University; Chief, Division of Rheumatology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee [373]

Jennifer M. Croswell, MD, MPH

Senior Program Officer, Office of the Chief Science Officer, Patient-Centered Outcomes Research Institute (PCORI), Washington, DC [70]

James E. Crowe, Jr., MD

Director, Vanderbilt Vaccine Center; Ann Scott Carell Chair and Professor, Departments of Pediatrics, Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee [199]

Ian Crozier, MD

NIH/NIAID/DCR Integrated Research Facility at Fort Detrick, Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, Frederick, Maryland [209, 210]

Philip E. Cryer, MD

Professor Emeritus of Medicine, Division of Endocrinology, Metabolism & Lipid Research, Washington University School of Medicine in St. Louis, St. Louis, Missouri [406]

Gary C. Curhan, MD

Professor of Medicine, Harvard Medical School; Professor of Epidemiology, Harvard School of Public Health; Channing Division of Network Medicine/Renal Division, Brigham and Women's Hospital, Boston, Massachusetts [318]

Brendan D. Curti, MD

Director, Cytokine and Adoptive Immunotherapy; Director, Genitourinary Oncology Research and Director, Melanoma Program; Robert W. Franz Endowed Chair for Clinical Research, Earle A. Chiles Research Institute, a Division of the Providence Cancer Institute, Portland, Oregon [76]

John J. Cush, MD

Executive Editor, *RheumNow.com*; Professor of Internal Medicine, University of Texas Southwestern Medical School, Dallas, Texas [370]

Charles A. Czeisler, MD, PhD

Frank Baldino, Jr, PhD Professor of Sleep Medicine, Professor of Medicine and Director, Division of Sleep Medicine, Harvard Medical School; Chief, Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Massachusetts [31]

Carolyn M. D'Ambrosio, MS, MD

Associate Professor of Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [39]

Josep Dalmau, MD, PhD

ICREA Professor, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Hospital Clínic, University of Barcelona, Barcelona, Spain; Adjunct Professor, University of Pennsylvania, Philadelphia, Pennsylvania [94]

Inger K. Damon, MD, PhD

Director, Division of High-Consequence Pathogens and Pathology (DHCPP), Centers for Disease Control and Prevention, Atlanta, Georgia [196]

Daniel F. Danzl, MD

Professor and Emeritus Chair, Department of Emergency Medicine, University of Louisville, Louisville, Kentucky [464, 465]

Robert B. Daroff, MD

Professor and Chair Emeritus, Department of Neurology, Case Western Reserve University School of Medicine; University Hospitals–Cleveland Medical Center, Cleveland, Ohio [22]

Jaideep Das Gupta, MD

Vascular Surgery Fellow, Vascular Surgery Division, University of California, San Diego, La Jolla, California [329]

Stephen N. Davis, MBBS, FRCP, FACE, MACP

Theodore E. Woodward Professor of Medicine; Professor of Physiology; Chairman, Department of Medicine, University of Maryland School of Medicine; Director, Institute for Clinical and Translational Research; Vice President of Clinical Translational Science
University of Maryland, Baltimore; Physician-in-Chief, University of Maryland Medical Center, Baltimore, Maryland [406]

Lisa M. DeAngelis, MD

Professor of Neurology, Weill Cornell Medical College; Physician-in-Chief and Chief Medical Officer, Memorial Sloan Kettering Cancer Center, New York, New York [90]

Rafael de Cabo, PhD

Chief, Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, Baltimore, Maryland [476]

Lucia De Franceschi, MD

Department of Medicine, University of Verona and AOUI Verona, Verona, Italy [100]

John Del Valle, MD

Professor and Vice Chair of Medicine, Department of Internal Medicine, University of Michigan School of Medicine, Ann Arbor, Michigan [324]

David W. Denning, MBBS, FRCP, FRCPath, FMedSci

Professor of Infectious Diseases in Global Health, The University of Manchester, Manchester, United Kingdom [217]

Akshay S. Desai, MD, MPH

Associated Professor of Medicine, Harvard Medical School; Director, Cardiomyopathy and Heart Failure, Advanced Heart Disease Section, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts [258]

Robert J. Desnick, PhD, MD

Dean for Genetic and Genomic Medicine Emeritus; Professor and Chairman Emeritus, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, Mount Sinai Health System, New York, New York [416]

Betty Diamond, MD

The Feinstein Institutes for Medical Research, Northwell Health System; Center for Autoimmunity and Musculoskeletal Diseases, Manhasset, New York [355]

Marcelo F. Di Carli, MD

Professor, Department of Radiology and Medicine, Harvard Medical School; Chief, Division of Nuclear Medicine and Molecular Imaging; Executive Director, Noninvasive Cardiovascular Imaging Program, Brigham and Women's Hospital, Boston, Massachusetts [241, A9]

Anna Mae Diehl, MD

Florence McAlister Professor of Medicine; Director, Duke Liver Center, Duke University, Durham, North Carolina [343]

Jules L. Dienstag, MD

Carl W. Walter Professor of Medicine, Harvard Medical School; Physician, Gastrointestinal Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts [339–341, 345, A13]

William P. Dillon, MD

Professor and Executive Vice-Chair, Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, California [423, A16]

Charles A. Dinarello, MD

Distinguished Professor of Medicine, University of Colorado School of Medicine, Aurora, Colorado [18]

R. Christopher Doiron, MD, FRCS(C)

Assistant Professor, Queens University at Kingston Canada; Staff Urologist, Department of Urology, Kingston Health Sciences Centre, Kingston, Ontario, Canada [51]

Anuja Dokras, MD, PhD

Professor of Obstetrics and Gynecology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [392, 393, 396]

Susan M. Domchek, MD

Bass Professor of Oncology, Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [467]

Richard L. Doty, PhD

Professor, Department of Otorhinolaryngology: Head and Neck Surgery; Director, Smell and Taste Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [33]

Vanja C. Douglas, MD

Professor of Neurology and Sara and Evan Williams Foundation Endowed Neurohospitalist Chair, University of California, San Francisco, San Francisco, California [23, 441]

David F. Driscoll, PhD

Associate Professor of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts [335]

Thomas D. DuBose, Jr., MD, MACP

Emeritus Professor of Medicine, Wake Forest School of Medicine, Winston Salem, North Carolina [55, S1]

J. Stephen Dumler, MD

Professor and Chairman, Department of Pathology, Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, Joint Pathology Center, Bethesda, Maryland [187]

Andrea Dunaif, MD

Lillian and Henry M. Stratton Professor of Molecular Medicine, System Chief, Hilda and J. Lester Gabrilove Division of Endocrinology, Diabetes and Bone Disease, Icahn School of Medicine and Mount Sinai Health System, New York, New York [398]

Samuel C. Durso, MD, MBA

Mason F. Lord Professor of Medicine; Executive Vice Chair, Johns Hopkins University Department of Medicine; Director, Department of Medicine, Johns Hopkins Bayview Medical Center, Baltimore, Maryland [36, A3]

Janice P. Dutcher, MD

Associate Director, Cancer Research Foundation of New York, Chappaqua, New York; Former Professor of Medicine, New York Medical College, Valhalla, New York [75]

Johanna T. Dwyer, DSc, RD

Professor of Medicine and Community Health, Tufts Medical School; Senior Nutrition Scientist (Contractor), Office of Dietary Supplements, National Institutes of Health, Boston, Massachusetts [332]

Kim A. Eagle, MD

Albion Walter Hewett Professor of Internal Medicine; Professor of Health Management Policy, School of Public Health; Director, Samuel and Jean Frankel Cardiovascular Center, University of Michigan, Ann Arbor, Michigan [480]

Sarah Rae Easter, MD

Assistant Professor of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School; Director of Obstetric Critical Care, Brigham and Women's Hospital, Boston, Massachusetts [479]

James A. Eastham, MD

Chief, Urology Service; Peter T. Scardino Chair in Oncology, Department of Surgery, Sidney Kimmel Center for Prostate and Urologic Cancers, Memorial Sloan Kettering Cancer Center, New York, New York [87]

Robert H. Eckel, MD

Professor of Medicine, Emeritus; University of Colorado School of Medicine, Aurora, Colorado [408]

John E. Edwards, Jr., MD

Distinguished Professor of Medicine Emeritus, David Geffen School of Medicine, University of California, Los Angeles; Senior Investigator, The Lundquist Institute and Emeritus Chief, Division of Infectious Disease at Harbor-UCLA Medical Center, Torrance, California [211, 216]

David A. Ehrmann, MD

Professor of Medicine, Section of Endocrinology; Director, University of Chicago Center for PCOS, University of Chicago, Chicago, Illinois [394]

Ramy H. Elshaboury, PharmD

Clinical Pharmacy Manager; Director, PGY2 Infectious Diseases Pharmacy Residency, Massachusetts General Hospital, Boston, Massachusetts [144]

Ezekiel J. Emanuel, MD, PhD

Chair, Department of Medical Ethics and Health Policy, Levy University Professor, Perelman School of Medicine and Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania [12]

Jack Ende, MD

The Schaeffer Professor of Medicine; Assistant Vice President, University of Pennsylvania Health System; Assistant Dean for Advanced Medical Practice, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, Pennsylvania [478]

John W. Engstrom, MD

Betty Anker Fife Distinguished Professor and Vice-Chairman; Neurology Residency Program Director, University of California, San Francisco, San Francisco, California [17, 440]

Moshe Ephros, MD

Clinical Associate Professor of Pediatrics, Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel [172]

Aaron C. Ermel, MD

Assistant Professor of Clinical Medicine, Department of Internal Medicine, Division of Infectious Diseases, Indiana University School of Medicine, Indianapolis, Indiana [198]

Tim Evans, DPhil, MD

Associate Dean and Director, School of Population and Global Health, McGill University, Montreal, Quebec, Canada [474]

Carmella Evans-Molina, MD, PhD

Eli Lilly Professor of Pediatric Diabetes; Professor, Departments of Pediatrics and Medicine; Director of the Center for Diabetes and Metabolic Diseases; Director of Diabetes Research, Herman B. Wells Center for Pediatric Research; Indiana University School of Medicine; Staff Physician, Richard L. Roudebush VA Medical Center, Indianapolis, Indiana [403]

Christopher H. Fanta, MD

Professor of Medicine, Harvard Medical School; Member, Pulmonary and Critical Care Medicine Division, Brigham and Women's Hospital; Director, Partners Asthma Center, Boston, Massachusetts [38]

Paul E. Farmer, MD, PhD

Kolokotrones University Professor, Harvard University; Chair, Department of Global Health and Social Medicine, Harvard Medical School; Chief, Division of Global Health Equity, Brigham and Women's Hospital; Chief Strategist, Co-Founder, Partners In Health, Boston, Massachusetts [472, 475]

I. Sadaf Farooqi, PhD, FRCP, FMedSci, FRS

Professor of Metabolism and Medicine, Wellcome-MRC Institute of Metabolic Science, University of Cambridge, Cambridge, United Kingdom [401]

Anthony S. Fauci, MD

Chief, Laboratory of Immunoregulation; Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [1, 5, 201, 202, 349, 350, 363, A14, S3]

David P. Faxon, MD

Senior Lecturer, Harvard Medical School; Associate Chief, Cardiovascular Medicine, Department of Medicine; Brigham and Women's Hospital, Boston, Massachusetts [242, 276, A11]

David Feller-Kopman, MD

Professor of Medicine, Dartmouth Geisel School of Medicine; Section Chief, Pulmonary and Critical Care Medicine, Dartmouth Hitchcock Medical Center, Lebanon, New Hampshire [299]

David T. Felson, MD, MPH

Professor of Medicine and Epidemiology, Boston University School of Medicine, Boston, Massachusetts [371]

Howard L. Fields, MD, PhD

Professor, Department of Neurology, University of California, San Francisco, San Francisco, California [13]

Gregory A. Filice, MD

Staff Physician, Veterans Affairs Medical Center; Professor of Medicine and Adjunct Professor of Public Health, University of Minnesota, Minneapolis, Minnesota [174]

Robert W. Finberg, MD[‡]

Richard M. Haidack Distinguished Professor of Medicine; Professor, Microbiology and Physiological Systems; Chair, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, Massachusetts [74, 143]

Courtney Finlayson, MD

Associate Professor, Division of Endocrinology, Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, Illinois [390]

Alain Fischer, MD, PhD

Imagine Institute; Professor at College de France, Paris, France [351, S8]

Erik Fisher, MD

Clinical Assistant Professor of Emergency Medicine, University of South Carolina School of Medicine Greenville; Director, Medical Toxicology, Department of Emergency Medicine, Prisma Health-Update, Greenville, South Carolina [460]

Agnes B. Fogo, MD

John L. Shapiro Endowed Chair in Pathology; Professor of Pathology, Medicine and Pediatrics; Director, Renal Pathology/Electron Microscopy Laboratory, Vanderbilt University Medical Center, Nashville, Tennessee [A4]

[‡]Deceased

Gregory K. Folkers, MS, MPH

Chief of Staff, Office of the Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [202]

Michael J. Fowler, MD

Associate Professor of Medicine, Division of Diabetes, Endocrinology and Metabolism, Department of Medicine; Course Director, Physical Diagnosis; Medical Director, Glucose Management Service; Director of Clinical Skills Development in Undergraduate Medical Education, Vanderbilt University School of Medicine, Nashville, Tennessee [404]

David M. Frazer, PhD

Associate Professor, Molecular Nutrition Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia [414]

Jane E. Freedman, MD

Director, Division of Cardiology, Physician-in-Chief, Vanderbilt Medical Center, Nashville, Tennessee [117]

Roy Freeman, MD

Professor of Neurology, Harvard Medical School; Director, Center for Autonomic and Peripheral Nerve Disorders, Beth Israel Deaconess Medical Center, Boston, Massachusetts [21]

Lawrence S. Friedman, MD

Professor of Medicine, Harvard Medical School; Professor of Medicine, Tufts University School of Medicine; The Anton R. Fried, MD Chair, Department of Medicine, Newton-Wellesley Hospital, Newton, Massachusetts; Assistant Chief of Medicine, Massachusetts General Hospital, Boston, Massachusetts [50]

Sonia Friedman, MD

Associate Professor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [326]

Andre D. Furtado, MD

Assistant Professor, Department of Radiology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania [A16]

Nicholas B. Galifianakis, MD, MPH

Associate Professor of Neurology, Movement Disorder and Neuromodulation Center Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, California [V1]

John I. Gallin, MD

Associate Director for Clinical Research; Chief Scientific Officer, Clinical Center, National Institutes of Health, Bethesda, Maryland [64]

Karunesh Ganguly, MD, PhD

Department of Neurology, University of California, San Francisco; Neurology & Rehabilitation Service, San Francisco VA Medical Center, San Francisco, California [487]

Olivier Garraud, MD, PhD

Professor, INSERM 1059, University of Lyon, Faculty of Medicine of Saint-Etienne, Saint-Etienne, France [113]

Gregory M. Gauthier, MD

Associate Professor, Department of Medicine, Division of Infectious Disease, School of Medicine and Public Health, University of Wisconsin, Madison, Madison, Wisconsin [214]

Charlotte A. Gaydos, MS, MPH, DrPH

Professor of Medicine, Division of Infectious Diseases, Johns Hopkins University, Baltimore, Maryland [189]

J. Michael Gaziano, MD, MPH

Professor of Medicine, Harvard Medical School; Physician, Brigham and Women's Hospital and the VA Boston Healthcare System, Boston, Massachusetts [238]

Thomas A. Gaziano, MD, MSc

Associate Professor of Medicine, Harvard Medical School; Associate Professor, Health Policy and Management, Center for Health Decision Sciences, Harvard School of Public Health; Director, Global Cardiovascular Health Policy and Prevention Unit, Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [238]

Susan L. Gearhart, MD

Associate Professor, Surgery, Johns Hopkins Medical Institutions, Baltimore, Maryland [328]

Jeffrey A. Gelfand, MD

Professor of Medicine (Part-Time), Harvard Medical School; Attending Physician, Infectious Diseases Division, Massachusetts General Hospital, Boston, Massachusetts [225]

Jeffrey M. Gelfand, MD, MAS, FAAN

Associate Professor of Neurology, Department of Neurology, University of California, San Francisco, San Francisco, California [23]

Lianne S. Gensler, MD

Professor of Medicine; Rheumatology Fellowship Program Director; Director, Spondyloarthritis Research Program and Clinic, University of California, San Francisco, San Francisco, California [362]

Alfred L. George, Jr., MD

Magerstadt Professor and Chair, Department of Pharmacology, Northwestern University Feinberg School of Medicine, Chicago, Illinois [309]

Dale N. Gerding, MD

Research Physician, Edward Hines Jr. VA Hospital, Hines, Illinois; Professor of Medicine (Retired), Loyola University Chicago Stritch School of Medicine, Maywood, Illinois [134]

Michael D. Geschwind, MD, PhD

Professor of Neurology; Michael J. Homer Chair in Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [438, V1]

Marc G. Ghany, MD, MHS

Tenure-Track Investigator, Liver Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [336]

Lorenzo Giacani, PhD

Associate Professor of Medicine, Department of Medicine, Division of Allergy & Infectious Diseases, University of Washington, Seattle, Washington [183]

Matthew P. Giannetti, MD

Division of Allergy and Clinical Immunology, Brigham and Women's Hospital; Harvard Medical School, Boston, Massachusetts [354]

Michael Giladi, MD, MSc

Associate Professor of Medicine, Sackler Faculty of Medicine, Tel Aviv University; Senior Physician, The Infectious Disease Unit; Director, The Bernard Pridan Laboratory for Molecular Biology of Infectious Diseases, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel [172]

Robert P. Giugliano, MD, SM, FACC, FAHA

Professor of Medicine, Harvard Medical School; Senior Investigator, TIMI Study Group, Cardiovascular Medicine, Brigham and Women's Hospital, Boston, Massachusetts [274]

Michael M. Givertz, MD

Professor of Medicine, Harvard Medical School; Medical Director, Heart Transplant and Mechanical Circulatory Support, Brigham and Women's Hospital, Boston, Massachusetts [257]

Roger I. Glass, MD, PhD

Director, Fogarty International Center; Associate Director for International Research National Institutes of Health, Bethesda, Maryland [203]

Seth R. Glassman, MD

Assistant Clinical Professor of Medicine, Division of Infectious Diseases, University at Buffalo Jacobs School of Medicine and the Biomedical Sciences, Buffalo, New York [176]

Eli Glatstein, MD[‡]

Professor Emeritus, Department of Radiation Oncology, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania [S5]

Ronald S. Go, MD

Chair, Core/Consultative Hematology, Division of Hematology, Mayo Clinic Rochester, Rochester, Minnesota [317]

[‡]Deceased

Peter J. Goadsby, MD, PhD, DSc, FRACP, FRCP, FMedSci
 Professor, NIHR-Wellcome Trust King's Clinical Research Facility, King's College London, United Kingdom; Professor, Department of Neurology, University of California, Los Angeles, Los Angeles, California [16, 430]

Hilary J. Goldberg, MD, MPH
 Assistant Professor of Medicine, Harvard Medical School; Medical Director, Lung Transplant Program; Clinical Director, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [286, 298]

Marcia B. Goldberg, MD
 Professor of Medicine and Microbiology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts [120]

Ary L. Goldberger, MD
 Professor of Medicine, Harvard Medical School & Wyss Institute for Biotechnology Inspired Engineering at Harvard University; Director, Margaret and H.A. Rey Institute for Nonlinear Dynamics in Medicine; Associate Chief, Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center, Boston, Massachusetts [240, A7, A8]

David Goldblatt, MB, ChB, PhD
 Professor of Vaccinology and Immunology, University College London Institute of Child Health, London, United Kingdom [146]

Samuel Z. Goldhaber, MD
 Professor of Medicine, Harvard Medical School; Associate Chief and Clinical Director, Division of Cardiovascular Medicine; Director, Thrombosis Research Group, Brigham and Women's Hospital, Boston, Massachusetts [279]

Andrea Gori, MD
 Full Professor of Infectious Diseases, School of Medicine and Surgery, Department of Pathophysiology and Transplantation; Co-Director, Centre for Multidisciplinary Research in Health Science (MACH), University of Milan; Director, Infectious Diseases Unit, Department of Internal Medicine, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy [178]

Marga G.A. Goris, PhD, MSC
 Head OIE and National Collaborating Centre for Reference and Research on Leptospirosis, Department of Medical Microbiology, Amsterdam University Medical Centers, Amsterdam, The Netherlands [184]

Maria Luisa Gorno-Tempini, MD, PhD
 Professor, Department of Neurology; Language Neurobiology Lab, Memory and Aging Center; Dyslexia Center, University of California, San Francisco, San Francisco, California [V2]

Daniel J. Gottlieb, MD, MPH
 Associate Professor of Medicine, Harvard Medical School; Director, Sleep Disorders Center, VA Boston Healthcare System, Sleep Medicine Division, Brigham and Women's Hospital, Boston, Massachusetts [297]

Peter A. Gottlieb, MD
 Professor of Pediatrics and Medicine, Barbara Davis Center for Childhood Diabetes, University of Colorado School of Medicine, Aurora, Colorado [389]

Gregory A. Grabowski, MD
 Professor Emeritus, University of Cincinnati College of Medicine; Departments of Pediatrics, and Molecular Genetics, Biochemistry and Microbiology, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio [418]

Yonatan H. Grad, MD, PhD
 Associate Professor of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts [121]

Christine Grady, RN, PhD
 Chief, Department of Bioethics, National Institutes of Health Clinical Center, Bethesda, Maryland [11]

Francesc Graus, MD, PhD
 Neuroimmunology Program, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic, Barcelona, Spain [94]

Steven A. Greenberg, MD
 Professor of Neurology, Harvard Medical School; Associate Neurologist, Brigham and Women's Hospital, Boston, Massachusetts [365]

Steven M. Greenberg, MD, PhD
 Professor of Neurology, Harvard Medical School; Vice Chair of Neurology, Massachusetts General Hospital, MGH Stroke Research Center, Boston, Massachusetts [433]

Norton J. Greenberger, MD[‡]
 Clinical Professor of Medicine, Harvard Medical School; Senior Physician, Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts [346]

Daryl R. Gress, MD
 Professor of Neurological Sciences; Director of Neurocritical Care, University of Nebraska Medical Center, Omaha, Nebraska [307, 429]

Patricia M. Griffin, MD
 Chief, Enteric Diseases Epidemiology Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [151]

Rasim Gucalp, MD, FACP
 Professor of Medicine, Albert Einstein College of Medicine; Associate Chairman for Educational Programs, Department of Oncology; Director, Hematology/Oncology Fellowship, Montefiore Medical Center, Bronx, New York [75]

Kalpana Gupta, MD, MPH
 Associate Chief of Staff and Chief, Infectious Diseases, Veterans Affairs Boston Healthcare System, West Roxbury, Massachusetts; Professor of Medicine, Boston University School of Medicine, Boston, Massachusetts [135]

Chadi A. Hage, MD
 Associate Professor of Clinical Medicine, Indiana University School of Medicine, Pulmonary Critical Care Medicine, Indianapolis, Indiana [212]

Bevra Hannahs Hahn, MD
 Distinguished Professor of Medicine (Emeritus), University of California, Los Angeles, Los Angeles, California [356]

Noah M. Hahn, MD
 Associate Professor of Oncology and Urology, Johns Hopkins University School of Medicine; Johns Hopkins University Greenberg Bladder Cancer Institute, Baltimore, Maryland [86]

Colin N. Haile, MD, PhD
 Assistant Professor, Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine; Michael E. DeBakey VA Medical Center, Houston, Texas [456]

Janet E. Hall, MD
 Clinical Director and Senior Investigator, Division of Intramural Research, NIH/NIEHS, Research Triangle Park, North Carolina [392, 393, 396]

Scott A. Halperin, MD
 Professor of Pediatrics and Microbiology & Immunology, Dalhousie University, Halifax, Nova Scotia, Canada [160]

Aidan Hampson, PhD
 Program and Scientific Officer, Special Content Expert on Cannabis, Clinical Research Grants Branch, Division of Therapeutics & Medical Consequences, National Institute on Drug Abuse, National Institutes of Health, Rockville, Maryland [455]

R. Doug Hardy, MD
 ID Specialists, Dallas, Texas [188]

Nigel Haroon, MD, PhD, DM, FRCPC
 Associate Professor of Medicine and Rheumatology, University of Toronto; Clinician Scientist and Attending Physician, University Health Network and Mount Sinai Hospital; Scientist, Krembil Research Institute, Toronto, Ontario, Canada [362]

[‡]Deceased

Phil A. Hart, MD

Associate Professor of Medicine; Director, Section of Pancreatic Disorders, Division of Gastroenterology, Hepatology, and Nutrition, The Ohio State University Wexner Medical Center, Columbus, Ohio [347, 348]

William L. Hasler, MD

Professor, Division of Gastroenterology and Hepatology, University of Michigan Health System, Ann Arbor, Michigan [45, 321]

Stephen L. Hauser, MD

Robert A. Fishman Distinguished Professor, Department of Neurology, University of California, San Francisco; Director, UCSF Weill Institute for Neurosciences, San Francisco, California [1, 5, 24, 25, 28, 422, 424, 441, 442, 444, 445, 447, S9]

Thomas R. Hawn, MD, PhD

Professor, Department of Medicine, Division of Allergy & Infectious Diseases, University of Washington, Seattle, Washington [159]

Daniel F. Hayes, MD, FASCO, FACP

Stuart B. Padnos Professor of Breast Cancer Research, University of Michigan Rogel Cancer Center, Ann Arbor, Michigan [79]

Barton F. Haynes, MD

Director, Duke Human Vaccine Institute; Frederic M. Hanes Professor of Medicine; Professor of Immunology, Departments of Medicine and Immunology, Duke University Medical Center, Durham, North Carolina [349, 350]

J. Claude Hemphill, III, MD, MAS

Professor of Neurology and Neurological Surgery, University of California, San Francisco; Chief, Neurology Service, Zuckerberg San Francisco General Hospital, San Francisco, California [307, 426–429]

Dirk M. Hentschel, MD

Assistant Professor of Medicine, Harvard Medical School; Director of Interventional Nephrology, Brigham Health; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [320]

Katherine A. High, MD

Professor Emerita, Perelman School of Medicine of the University of Pennsylvania; President, Therapeutics, Asklepios BioPharmaceuticals, Philadelphia, Pennsylvania [470]

Christine E. Hill-Kayser, MD

Assistant Professor of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [S5]

Ikuro Hirano, MD

Professor of Medicine, Division of Gastroenterology, Northwestern University Feinberg School of Medicine, Chicago, Illinois [44, 323]

Martin S. Hirsch, MD

Professor of Medicine, Harvard Medical School; Senior Physician, Massachusetts General Hospital, Boston, Massachusetts [195]

Dieter Hoelzer, PhD, MD

Emeritus Director of Internal Medicine, University of Frankfurt, Frankfurt, Germany [106]

A. Victor Hoffbrand, DM

Emeritus Professor of Haematology, University College, London, United Kingdom [99]

L. John Hoffer, MD, PhD

Professor, Faculty of Medicine, McGill University; Senior Physician, Divisions of Internal Medicine and Endocrinology, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada [335]

Charles W. Hoge, MD

Senior Scientist, Center for Psychiatry and Neuroscience, Walter Reed Army Institute of Research, Silver Spring, Maryland [S7]

Steven M. Holland, MD

Scientific Director, National Institute of Allergy and Infectious Diseases; Distinguished NIH Investigator, National Institutes of Health, Bethesda, Maryland [64, 180]

King K. Holmes, MD, PhD

Professor Emeritus, Global Health; Professor Emeritus, Medicine – Allergy and Infectious Diseases; Director, Research and Faculty Development, Department of Global Health; Co-Director, Center for AIDS Research, University of Washington – Fred Hutchinson Cancer Research Center; PI, International Training and Education Center for Health (I-TECH), University of Washington – University of California San Francisco; Director, Center for AIDS and STD, University of Washington; Infectious Disease Section Head, Harborview Medical Center; Member, National Institutes of Health (NIH) Fogarty International Center Council; Member, National Institutes of Health (NIH) Council or Councils, Seattle, Washington [136]

David Hong, MD

Assistant Professor of Medicine, Harvard Medical School; Division of Allergy & Immunology, Brigham and Women's Hospital, Boston, Massachusetts [353]

Jay H. Hoofnagle, MD

Director, Liver Diseases Research Branch, Division of Digestive Diseases and Nutrition, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [336]

David C. Hooper, MD

Professor of Medicine, Harvard Medical School; Chief, Infection Control Unit, and Associate Chief, Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts [144, 145]

Robert J. Hopkin, MD

Associate Professor of Clinical Pediatrics, Cincinnati Children's Hospital Medical Center Division of Human Genetics, Cincinnati, Ohio [418]

Leora Horn, MD, MSc

Associate Professor, Division of Hematology and Medical Oncology, Vanderbilt University School of Medicine, Nashville, Tennessee [78]

Jonathan C. Horton, MD, PhD

William F. Hoyt Professor of Neuro-ophthalmology, Professor of Ophthalmology, Neurology and Physiology, University of California, San Francisco School of Medicine, San Francisco, California [32, V3]

Howard Hu, MD, MPH, ScD

Professor & Flora L. Thornton Chair, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California [458]

Deborah T. Hung, MD, PhD

Professor of Genetics, Harvard Medical School, Brigham and Women's Hospital and Massachusetts General Hospital, Boston, Massachusetts; Co-Director, Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts [121]

Gary M. Hunninghake, MD, MPH

Associate Professor of Medicine, Harvard Medical School; Division of Pulmonary & Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [293]

Stephen C. Hunt, MD, MPH

National Director, VA Post-Deployment Integrated Care Initiative; Clinical Professor of Medicine, Department of Medicine, Division of General Internal Medicine, Occupational and Environmental Medicine Program, University of Washington, Seattle, Washington [S7]

Wade T. Iams, MD, MSCI

Department of Medicine, Division of Hematology/Oncology, Vanderbilt University Medical Center, Nashville, Tennessee [78]

Ashraf S. Ibrahim, PhD

Professor of Medicine, Division of Infectious Diseases, David Geffen School of Medicine at the University of California, Los Angeles; Senior Investigator and Vice Chair, Board of Directors, Director of the Graduate Studies Program, The Lindquist Institute at Harbor-UCLA Medical Center, Torrance, California [216, 218]

David H. Ingbar, MD

Professor, Medicine, Pediatrics and Integrative Biology and Physiology; Director, Pulmonary, Allergy, Critical Care and Sleep Division; CTSI Associate Director, Education, Career Development and Training; Executive Director, Center for Lung Science and Health, University of Minnesota, Minneapolis, Minnesota [305]

Elliot Israel, MD

Professor of Medicine, Harvard Medical School; Gloria M. and Anthony C. Simboli Distinguished Chair in Asthma Research; Director of Clinical Research, Pulmonary and Critical Care Division, Allergy and Immunology Division, Brigham and Women's Hospital, Boston, Massachusetts [287]

Elias Jabbour, MD

Professor, Section Chief, Acute Lymphocytic Leukemia, Department of Leukemia, Division of Cancer Medicine, MD Anderson Cancer Center, Houston, Texas [105]

Alan C. Jackson, MD

Professor of Medicine (Neurology), University of Manitoba, Winnipeg, Manitoba, Canada [208]

Yves Jackson, MD, PhD

Assistant Professor, Division of Primary Care Medicine, Geneva University Hospitals, Geneva, Switzerland [227]

Danny O. Jacobs, MD, MPH, FACS

President, Oregon Health and Science University, Portland, Oregon [15, 330, 331]

Caron A. Jacobson, MD

Assistant Professor of Medicine, Harvard Medical School; Dana-Farber Cancer Institute, Boston, Massachusetts [108, 109]

Rajesh K. Jain, MD

Assistant Professor, Department of Medicine, Section of Endocrinology, University of Chicago, Chicago, Illinois [412]

J. Larry Jameson, MD, PhD

Robert G. Dunlop Professor of Medicine; Dean, Raymond and Ruth Perelman School of Medicine; Executive Vice President, University of Pennsylvania for the Health System, Philadelphia, Pennsylvania [1, 5, 47, 93, 376–380, 382–385, 390, 391, 466, 467, A15]

Gordon L. Jensen, MD, PhD

Senior Associate Dean for Research; Professor of Medicine and Nutrition, University of Vermont Larner College of Medicine, Burlington, Vermont [334]

Savio John, MD

Chief of Gastroenterology, State University of New York Upstate Medical University, Syracuse, New York [49]

James R. Johnson, MD

Professor of Medicine, Division of Infectious Diseases and International Medicine, University of Minnesota, Minneapolis, Minnesota [161]

Stuart Johnson, MD

Professor of Medicine, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois; Staff Physician, Edward Hines Jr. VA Hospital, Hines, Illinois [134]

S. Claiborne Johnston, MD, PhD

Professor of Neurology, Dell Medical School, University of Texas at Austin, Austin, Texas [426–428]

S. Andrew Josephson, MD

Professor and Chairman, Department of Neurology, University of California, San Francisco, San Francisco, California [27, 28, 307, 422, V4]

Sandeep S. Jubbal, MD

Assistant Professor of Medicine, Division of Infectious Diseases and Immunology, University of Massachusetts Chan Medical School, Worcester, Massachusetts [141]

Harald Jüppner, MD

Professor of Pediatrics, Endocrine Unit and Pediatric Nephrology Unit, Harvard Medical School; Massachusetts General Hospital, Boston, Massachusetts [410]

Joseph Kado, MBBS, DCH, MMed

University of Western Australia, Crawley, Western Australia; Clinical Research Officer, Telethon Kids Institute, Nedlands, Western Australia [359]

Peter J. Kahrilas, MD

Gilbert H. Marquardt Professor of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [44, 323]

Stephen G. Kaler, MD

CAPT, US Public Health Service (Ret); Professor of Pediatrics and Genetics, The Ohio State University College of Medicine; Principal Investigator, Center for Gene Therapy, Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus, Ohio [415]

Gail Kang, MD

Private Practice, Berkeley, California [V1]

Hagop Kantarjian, MD

Chairman, Leukemia Department; Professor of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, Texas [105]

Hemanta K. Kar, MBBS, MD, MAMS

Professor and Head, Department of Dermatology, STD and Leprosy, Kalinga Institute of Medical Sciences, Bhubaneswar, Odisha, India [179]

Adolf W. Karchmer, MD

Professor of Medicine, Harvard Medical School; Emeritus Chief, Division of Infectious Diseases, Beth Israel Deaconess Medical Center, Boston, Massachusetts [128]

Dennis L. Kasper, MD

William Ellery Channing Professor of Medicine and Professor of Immunology, Department of Immunology, Harvard Medical School; Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts [1, 5, 119, 132, 177, 471]

Daniel L. Kastner, MD, PhD

Scientific Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland [369]

Carol A. Kauffman, MD

Chief, Infectious Diseases Section, VA Ann Arbor Healthcare System; Professor of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan [219]

Elaine T. Kaye, MD

Assistant Professor of Dermatology, Harvard Medical School, Boston Children's Hospital, Boston, Massachusetts [19, A1]

Kenneth M. Kaye, MD

Professor of Medicine, Harvard Medical School; Senior Physician, Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts [19, A1]

John F. Keaney, Jr., MD

Professor of Medicine, Harvard Medical School; Chief, Division of Cardiovascular Medicine; Co-Executive Director, Heart and Vascular Center, Brigham and Women's Hospital, Boston, Massachusetts [237]

David Kelsen, MD

Professor of Medicine, Weill Cornell Medical College; Edward S. Gordon Chair in Medical Oncology, Memorial Sloan Kettering Cancer Center, New York, New York [80]

John A. Kessler, MD

Davee Professor of Stem Cell Biology, Davee Department of Neurology; Director, Northwestern University Stem Cell Institute, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [484]

Maryam Ali Khan, MD

Research Scholar, Division of Vascular and Endovascular Surgery, University of California, San Diego, San Diego, California [329]

Sundeept Khosla, MD

Dr. Francis Chuker and Nathan Landow Research Professor; Mayo Foundation Distinguished Investigator, Mayo Clinic College of Medicine, Rochester, Minnesota [54]

Kiran K. Khush, MD, MAS

Professor of Medicine (Cardiovascular Medicine), Stanford University School of Medicine, Stanford, California [490]

Anthony A. Killeen, MD, PhD

Professor, Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota [S10]

Kami Kim, MD

Andor Szentivanyi Professor of Medicine; Director, Division of Infectious Diseases and International Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida [228]

Priya S. Kishnani, MD

C.L. and Su Chen Professor of Pediatrics; Medical Director, YT and Alice Chen Pediatrics Genetics and Genomics Center; Division Chief, Medical Genetics; Professor of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina [419]

Bruce S. Klein, MD

Gerard B. Odell Professor and Shirley S. Matchette Professor; Chief, Division of Pediatric Infectious Disease, Departments of Pediatrics, Medicine and Medical Microbiology and Immunology, University of Wisconsin, Madison, Madison, Wisconsin [214]

Christine Klein, MD

Professor of Neurology and Neurogenetics, Institute of Neurogenetics and Department of Neurology, University of Lübeck and University Hospital Schleswig-Holstein, Lübeck, Germany [436]

David M. Knipe, PhD

Higgins Professor of Microbiology and Molecular Genetics; Head, Program in Virology, Department of Microbiology, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts [190]

Isaac S. Kohane, MD, PhD

Marion V. Nelson Professor and Chair, Biomedical Informatics; Harvard Medical School; Faculty Member, Informatics Program, Boston Children's Hospital, Boston, Massachusetts [488]

Barbara A. Konkle, MD

Professor of Medicine/Hematology, University of Washington; Scientific Director, Washington Center for Bleeding Disorders, Seattle, Washington [65, 115]

Bruce A. Koplan, MD, MPH

Assistant Professor of Medicine, Harvard Medical School; Director, Electrophysiology Laboratory, Brigham and Women's Hospital, Boston, Massachusetts [243–245]

Peter Kopp, MD

Professor of Medicine/Médecin Chef, Division of Endocrinology, Diabetology and Metabolism, University of Lausanne, Lausanne, Switzerland; Adjunct Professor, Division of Endocrinology, Metabolism and Molecular Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [466]

Walter J. Koroshetz, MD

National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland [139]

Thomas R. Kosten, MD

J. H. Waggoner Professor of Psychiatry, Pharmacology, Immunology, Neuroscience, Baylor College of Medicine, Houston, Texas [456]

Theodore A. Kotchen, MD[#]

Professor Emeritus, Associate Dean for Clinical Research, Medical College of Wisconsin, Milwaukee, Wisconsin [277]

Camille Nelson Kotton, MD, FIDSA, FAST

Clinical Director, Transplant and Immunocompromised Host Infectious Diseases, Infectious Diseases Division, Massachusetts General Hospital, Boston, Massachusetts [195]

Barnett S. Kramer, MD, MPH, FACP

Director, Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland [70]

Joel Kramer, PsyD

John Douglas French Alzheimer's Foundation Endowed Professor of Neuropsychology in Neurology; Director of Neuropsychology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [V2]

Arnold R. Kriegstein, MD, PhD

Professor of Neurology, University of California, San Francisco, San Francisco, California [424]

Somashekhar G. Krishna, MD, MPH

Professor of Medicine, Division of Gastroenterology, Hepatology, & Nutrition, The Ohio State University Wexner Medical Center, Columbus, Ohio [347, 348]

Henry M. Kronenberg, MD

Professor of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts [409]

Jens H. Kuhn, MD, PhD, MS

Principal Scientist and Director of Virology, NIH/NIAID/DCR/Integrated Research Facility at Fort Detrick, Frederick, Maryland [209, 210]

Matthew H. Kulke, MD

Zoltan Kohn Professor of Medicine, Boston University School of Medicine; Chief, Section of Hematology and Medical Oncology, Boston Medical Center; Co-Director, Boston University–Boston Medical Cancer Center, Boston, Massachusetts [84]

Sebastian G. Kurz, MD

Associate Professor of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, The Mount Sinai Hospital, New York, New York [181]

Robert F. Kushner, MD

Professor of Medicine and Medical Education, Northwestern University Feinberg School of Medicine, Chicago, Illinois [402]

Raymond Y. Kwong, MD, MPH, FACC

Professor of Medicine, Harvard Medical School; Director of Cardiac Magnetic Resonance Imaging, Cardiovascular Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [241, A9]

Loren Laine, MD

Professor of Medicine; Chief, Section of Digestive Diseases, Yale School of Medicine, New Haven, Connecticut; VA Connecticut Healthcare System, West Haven, Connecticut [48]

Neal K. Lakdawala, MD, MSc

Assistant Professor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [259]

Anil K. Lalwani, MD

Associate Dean for Student Research, Columbia University Vagelos College of Physicians and Surgeons; Professor and Vice Chair for Research; Co-Director, Columbia Cochlear Implant Center, Columbia University Vagelos College of Physicians and Surgeons; Medical Director of Perioperative Services, New York Presbyterian–Columbia University Irving Medical Center, New York, New York [34]

Michael J. Landzberg, MD

Associate Professor of Medicine, Harvard Medical School; Boston Adult Congenital Heart Disease and Pulmonary Hypertension Program, Boston Children's Hospital, Brigham and Women's Hospital, Boston, Massachusetts [269]

H. Clifford Lane, MD

Clinical Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [202, S3]

Helene M. Langevin, MD

Director, National Center for Complementary and Integrative Health, National Institutes of Health, Bethesda, Maryland [482]

Carol A. Langford, MD, MHS

Harold C. Schott Endowed Chair; Director, Center for Vasculitis Care and Research, Department of Rheumatic and Immunologic Diseases, Cleveland Clinic, Cleveland, Ohio [363, 366, 374, 375, A14]

[#]Deceased

Regina C. LaRocque, MD, MPH

Associate Professor of Medicine, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts [133]

Leslie P. Lawley, MD

Associate Professor, Department of Dermatology, School of Medicine, Emory University, Atlanta, Georgia [57]

Thomas J. Lawley, MD

William Patterson Timmie Professor of Dermatology, Former Dean, Emory University School of Medicine, Atlanta, Georgia [56, 59, A5]

David G. Le Couteur, MD, PhD

Professor of Geriatric Medicine, University of Sydney; Senior Staff Geriatrician, Concord Hospital, Sydney, Australia [476]

Sancy A. Leachman, MD, PhD

John D. Gray Endowed Chair in Melanoma Research; Professor & Chair, Department of Dermatology, Oregon Health & Science University, Center for Health & Healing, Portland, Oregon [76]

William M. Lee, MD

Professor of Internal Medicine; Meredith Mosle Chair in Liver Diseases, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas [340]

Charles Lei, MD

Assistant Professor, Department of Emergency Medicine, Vanderbilt University Medical Center, Nashville, Tennessee [460]

Jane A. Leopold, MD

Associate Professor of Medicine, Harvard Medical School; Director, Women's Interventional Cardiology Health Initiative, Brigham and Women's Hospital, Boston, Massachusetts [242, A11]

Jessica Leung, MPH

Epidemiologist, Viral Vaccine Preventable Diseases Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [207]

Nelson Leung, MD

Professor of Medicine, Division of Nephrology and Hypertension, Division of Hematology, Mayo Clinic Rochester, Rochester, Minnesota [317]

Jonathan S. Leventhal, MD

Assistant Professor of Dermatology, Yale University School of Medicine, New Haven, Connecticut [58]

Bruce D. Levy, MD

Professor of Medicine, Harvard Medical School; Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [284, 301]

Julia B. Lewis, MD

Professor of Medicine, Division of Nephrology and Hypertension, Vanderbilt University Medical Center, Nashville, Tennessee [314]

Peter Libby, MD

Mallinckrodt Professor of Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [A10]

Richard W. Light, MD, FCCP[†]

Professor of Medicine, Division of Allergy, Pulmonary, and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, Tennessee [294, 295]

Jin-Mann S. Lin, PhD

Division of High-Consequence Pathogens and Pathology, National Center for Zoonotic and Emerging Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [450]

Jeffrey A. Linder, MD, MPH, FACP

Michael A. Gertz Professor of Medicine and Chief, Division of General Internal Medicine and Geriatrics, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [35]

Robert Lindsay, MD, PhD

Professor of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York; Chief, Internal Medicine; Attending Physician, Helen Hayes Hospital, West Haverstraw, New York [411]

Michail S. Lionakis, MD, ScD

Chief, Fungal Pathogenesis Section, Laboratory of Clinical Immunology & Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [211, 216]

Marc E. Lippman, MD, MACP, FRCP

Professor of Oncology and Internal Medicine, Georgetown University, Washington, DC [79]

Peter E. Lipsky, MD

Charlottesville, Virginia [355]

Irene Litvan, MD, MSc, FAAN, FANA

Tasch Endowed Professor in Parkinson Disease Research; Director of the Parkinson and Other Movement Disorders Center, University of California, San Diego, La Jolla, California [434]

Eva S. Liu, MD

Assistant Professor of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts [409]

Kathleen D. Liu, MD, PhD, MAS

Professor, Division of Nephrology, Department of Medicine, Division of Critical Care Medicine, Department of Anesthesiology, University of California, San Francisco, San Francisco, California [312]

Josep M. Llovet, MD, PhD

Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York; Liver Cancer Translational Research Laboratory, Barcelona Clínic Liver Cancer Group (BCLC), Liver Unit, IDIBAPS-Hospital Clínic, CIBERehd, University of Barcelona, Catalonia, Spain; Institutó Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain [82]

Donald M. Lloyd-Jones, MD, ScM, FACC, FAHA

Eileen M. Foell Professor of Heart Research; Professor of Preventive Medicine, Medicine, and Pediatrics; Chair, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine; President, American Heart Association 2021–22, Chicago, Illinois [2]

Bernard Lo, MD

Professor of Medicine Emeritus and Director Emeritus of the Program in Medical Ethics, University of California, San Francisco, San Francisco, California; President Emeritus, The Greenwall Foundation, New York, New York [11]

George Loewenstein, PhD

Herbert A. Simon Professor of Economics and Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania [481]

Dan L. Longo, MD

Professor of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital; Deputy Editor, *New England Journal of Medicine*, Boston, Massachusetts [1, 5, 62, 63, 66, 69, 72, 73, 93, 95, 96, 101, 108–111, 201, A6]

Nicola Longo, MD, PhD

Professor and Chief, Division of Medical Genetics, Departments of Pediatrics, Pathology, Nutrition, and Integrated Physiology; Medical Co-Director, Biochemical Genetics Laboratory, ARUP Laboratories, University of Utah, Salt Lake City, Utah [420, 421]

Lenny López, MD, MPH, MDiv

Professor of Medicine, University of California San Francisco; San Francisco VA Medical Center, San Francisco, California [10]

Joseph Loscalzo, MD, PhD

Hersey Professor of the Theory and Practice of Medicine, Harvard Medical School; Chairman, Department of Medicine, Soma Weiss MD Distinguished Chair in Medicine, Physician-in-Chief, Brigham and Women's Hospital, Boston, Massachusetts [1, 5, 40–43, 117, 236, 237, 239, 259, 261–268, 270, 273, 275, 280–283, 486, 492]

[†]Deceased

Daniel H. Lowenstein, MD

Dr. Robert B. and Mrs. Ellinor Aird Professor of Neurology; Executive Vice Chancellor and Provost, University of California, San Francisco, San Francisco, California [422, 425, V6]

Elyse E. Lower, MD

Department of Internal Medicine, Division of Hematology-Oncology, University of Cincinnati, Cincinnati, Ohio [367]

Franklin D. Lowy, MD

Clyde '56 and Helen Wu Professor Emeritus of Medicine and Professor Emeritus of Pathology and Cell Biology (in Epidemiology), Columbia University College of Physicians and Surgeons, New York, New York [147]

Sheila A. Lukehart, PhD

Professor of Medicine, Division of Allergy & Infectious Diseases and Global Health, University of Washington, Seattle, Washington [182, 183]

Carolina Lúquez, PhD

Team Lead, National Botulism and Enteric Toxins Team, Enteric Diseases Laboratory Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [153]

Lucio Luzzatto, MD, FRCP, FRCPath

Professor of Haematology, Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, Tanzania; Honorary Professor of Hematology, University of Florence, Firenze, Italy [100]

Calum A. MacRae, MD, PhD

Professor of Medicine, Harvard Medical School; Vice Chair for Scientific Innovation, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [237]

Lawrence C. Madoff, MD

Professor of Medicine, University of Massachusetts Chan Medical School, Worcester, Massachusetts; Medical Director, Bureau of Infectious Disease and Laboratory Sciences, Massachusetts Department of Public Health, Hinton State Laboratory Institute, Jamaica Plain, Massachusetts [130, 141]

Barry J. Make, MD

Co-Director, COPD Program; Professor, Department of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, National Jewish Health, University of Colorado Denver School of Medicine, Denver, Colorado [292]

Mahmoud Malas, MD, MHS, RPVI, FACS

Professor in Residence; Vice Chair of Surgery for Clinical Research; Chief Division Vascular and Endovascular Surgery, University of California, San Diego, Health System, La Jolla, California [329]

Fransiska Malfait, MD, PhD

Associate Professor, Center for Medical Genetics, Ghent University Hospital and Department for Biomolecular Medicine, Ghent University, Ghent, Belgium [413]

Hari R. Mallidi, MD

Associate Professor of Surgery, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [298]

Susan J. Mandel, MD, MPH

Professor of Medicine; Chief, Division of Endocrinology, Diabetes and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [382–385]

Brian F. Mandell, MD, PhD

Professor and Chairman of Medicine, Cleveland Clinic Lerner College of Medicine, Department of Rheumatic and Immunologic Disease, Cleveland Clinic, Cleveland, Ohio [374]

Lionel A. Mandell, MD, FRCPC

Professor Emeritus of Medicine, McMaster University, Hamilton, Ontario, Canada [126]

Geoffrey T. Manley, MD, PhD

Professor and Vice Chairman of Neurological Surgery, University of California, San Francisco; Chief of Neurosurgery, Zuckerberg San Francisco General Hospital and Trauma Center; Co-Director, Brain and Spinal Injury Center, University of California, San Francisco, San Francisco, California [443]

Arjun K. Manrai, PhD

Assistant Professor, Harvard Medical School; Computational Health Informatics Program, Boston Children's Hospital, Boston, Massachusetts [488]

JoAnn E. Manson, MD, DrPH

Professor of Medicine and the Michael and Lee Bell Professor of Women's Health, Harvard Medical School; Chief, Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts [395]

Joan C. Marini, MD, PhD

Senior Investigator; Head, Section on Heritable Disorders of Bone and Extracellular Matrix, National Institute of Child Health and Human Development (NICHD), National Institutes of Health, Bethesda, Maryland [413]

Daniel B. Mark, MD, MPH

Professor of Medicine, Duke University Medical Center; Director, Outcomes Research, Duke Clinical Research Institute, Durham, North Carolina [4]

Mariel Marlow, PhD, MPH

Mumps Program Lead, Viral Vaccine Preventable Diseases Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [207]

Alexander G. Marneros, MD, PhD

Associate Professor, Department of Dermatology, Harvard Medical School; Cutaneous Biology Research Center, Massachusetts General Hospital, Boston, Massachusetts [61]

Bradley A. Maron, MD

Associate Professor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [283]

Jeanne M. Marrazzo, MD, MPH

Professor of Medicine; Director, Division of Infectious Diseases, University of Alabama at Birmingham, Birmingham, Alabama [136]

Gary J. Martin, MD

Raymond J. Langenbach, MD Professor of Medicine; Senior Vice Chairman, Department of Medicine, Northwestern University Medical School, Chicago, Illinois [6]

Anthony F. Massaro, MD

Instructor, Harvard Medical School; Director, Medical Intensive Care Unit, Division of Pulmonary and Critical Care, Brigham and Women's Hospital, Boston, Massachusetts [300, 303]

Henry Masur, MD

Chief, Critical Care Medicine Department, National Institutes of Health Clinical Center, Bethesda, Maryland [220]

Max Maurin, MD, PhD

Professor of Bacteriology, Université Grenoble Alpes; Centre Hospitalier Universitaire, Institut de Biologie et Pathologie, Grenoble, France [170]

Marcela V. Maus, MD, PhD

Director, Cell Therapy Program; Paula O'Keefe Endowed Chair, Massachusetts General Hospital Cancer Center; Associate Professor of Medicine, Harvard Medical School; Attending Physician, Hematopoietic Cell Transplant & Cell Therapy Program, Massachusetts General Hospital; Associate Member, Broad Institute of MIT and Harvard; Associate Member, Ragon Institute of MGH, MIT, and Harvard, Charlestown, Massachusetts [470]

Clio P. Mavragani, MD

Rheumatologist, Associate Professor, Department of Physiology, National and Kapodistrian University of Athens, Athens, Greece [357, 361]

Robert J. Mayer, MD

Faculty Vice President for Academic Affairs, Dana-Farber Cancer Institute; Stephen B. Kay Family Professor of Medicine, Harvard Medical School, Boston, Massachusetts [81]

Jared R. Mayers, MD, PhD

Research Fellow in Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [489]

Sarah Mbaeyi, MD, MPH

Medical Officer, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [123]

Alexander J. McAdam, MD, PhD

Associate Professor of Pathology, Harvard Medical School; Medical Director, Infectious Diseases Diagnostic Laboratory, Boston Children's Hospital, Boston, Massachusetts [S11]

Calvin O. McCall, MD

Dermatology Section, Hunter Holmes McGuire Veterans Affairs Medical Center, Richmond, Virginia [A5]

Zachary B. R. McClain, MD

Attending Physician, Gender and Sexuality Development Clinic; Medical Director, Young Men's Clinic, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania [400]

John F. McConville, MD

Associate Professor of Medicine; Director, Internal Medicine Residency Program, Vice Chair for Education, University of Chicago, Chicago, Illinois [296]

Michael McCrea, PhD, ABPP

Professor and Eminent Scholar; Vice Chair of Research; Co-Director, Center for Neurotrauma Research (CNTR), Department of Neurosurgery, Medical College of Wisconsin, Milwaukee, Wisconsin [443]

Kathleen M. McKibbin, MD

Staff Physician, Northwestern University Health Services, Evanston, Illinois [2]

Maureen McMahon, MD, MCR

Associate Chief; Associate Professor, Division of Rheumatology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California [356]

Kevin T. McVary, MD, FACS

Director of the Center for Male Health; Professor of Urology, Department of Urology, Stritch School of Medicine, Loyola University Medical Center, Maywood, Illinois [397]

John N. Mecchella, DO, MPH

Assistant Professor of Medicine, Geisel School of Medicine at Dartmouth, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire [417]

Mandeep R. Mehra, MD, MSc, FRCP (London)

Professor of Medicine, Harvard Medical School; The William Harvey Distinguished Chair in Advanced Cardiovascular Medicine; Executive Director, Center for Advanced Heart Disease, Brigham and Women's Hospital, Boston, Massachusetts [257, 258, 260]

Sanjay R. Mehta, MD, DTM&H, D(ABMM)

Associate Professor of Medicine and Pathology, University of California, San Diego School of Medicine, San Diego, California [S12]

Shlomo Melmed, MBChB, MACP, FRCP

Executive Vice President and Dean of the Medical Faculty; Professor of Medicine, Cedars-Sinai Medical Center, Los Angeles, California [378–380]

Robert O. Messing, MD

Professor and Chair of Neuroscience; Professor of Neurology; Director, Waggoner Center for Alcohol and Addiction Research, University of Texas at Austin, Austin, Texas [451]

Nancy Messonnier, MD

Executive Director for Pandemic Prevention and Health Systems, Skoll Foundation, Palo Alto, California [123]

M.-Marsel Mesulam, MD

Ruth Dunbar Davee Professor of Neuroscience and Neurology, Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University Feinberg School of Medicine, Chicago, Illinois [30]

Robert G. Micheletti, MD

Associate Professor of Dermatology and Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [60]

Aaron W. Michels, MD

Associate Professor of Pediatrics, Medicine, and Immunology, Barbara Davis Center for Childhood Diabetes, University of Colorado School of Medicine, Aurora, Colorado [389]

Susan Miesfeldt, MD

Associate Professor of Medicine, Tufts University School of Medicine; Medical Oncology, Medical Director, Cancer Risk and Prevention Program, Maine Medical Center Cancer Institute, Scarborough, Maine [467]

Bruce L. Miller, MD

A. W. and Mary Margaret Clausen Distinguished Professor of Neurology, Memory and Aging Center, Global Brain Health Institute, University of California, San Francisco School of Medicine, San Francisco, California [27, 29, 431, 432, 434, V2]

Samuel I. Miller, MD

Professor of Medicine, Microbiology and Genome Sciences, University of Washington, Seattle, Washington [165]

William R. Miller, MD

Assistant Professor of Medicine, Division of Infectious Diseases, Houston Methodist Hospital, Center for Infectious Diseases Research, Houston Methodist Research Institute, Houston, Texas [149]

Jyoti Mishra, PhD

Department of Psychiatry, University of California, San Diego, La Jolla, California [487]

Hana Mitchell, MD, MSC

Clinical Assistant Professor, Division of Pediatric Infectious Diseases, Department of Pediatrics, The University of British Columbia, BC Children's Hospital, Vancouver, British Columbia, Canada [3]

Simon J. Mitchell, MBChB, PhD, FUHM, FANZCA

Professor, Department of Anaesthesiology, University of Auckland and Auckland City Hospital, Auckland, New Zealand [463]

Babak Mokhlesi, MD, MSc

The J. Bailey Carter, MD Professor of Medicine; Chief, Division of Pulmonary, Critical Care, and Sleep Medicine; Co-Director, Rush Lung Center, Rush University Medical Center, Chicago, Illinois [296]

Thomas A. Moore, MD, FACP, FIDSA

Clinical Professor of Medicine, University of Kansas School of Medicine-Wichita Campus, Wichita, Kansas [222]

Richard I. Morimoto, PhD

Bill and Gayle Cook Professor of Biology, Department of Molecular Biosciences, Rice Institute for Biomedical Research, Northwestern University, Evanston, Illinois [491]

Alison Morris, MD, MS

Chief, Pulmonary, Allergy and Critical Care Medicine; Professor of Medicine; UPMC Chair of Translational Pulmonary and Critical Care Medicine; Director, University of Pittsburgh Center for Medicine and the Microbiome, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania [220]

David A. Morrow, MD, MPH

Professor of Medicine, Harvard Medical School; Director, Samuel A. Levine Cardiac Intensive Care Unit, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts [14]

William J. Moss, MD, MPH

Professor, Departments of Epidemiology, International Health, and Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland [205]

Robert J. Motzer, MD

Jack and Dorothy Byrne Chair in Clinical Oncology, Kidney Cancer Section Head; Attending Physician, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York [85]

David B. Mount, MD, FRCPC

Assistant Professor of Medicine, Harvard Medical School; Clinical Chief and Director, Dialysis Services Renal Divisions, Brigham and Women's Hospital and VA Boston Healthcare System; Boston, Massachusetts [52, 53, S1]

Haralampos M. Moutsopoulos, MD, FACP, FRCP(hc), Master ACR
Professor, Chair Medical Sciences-Immunology, Academy of Athens, Athens, Greece [357, 361]

Catharina M. Mulders-Manders, MD, PhD

Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands [20]

L. Silvia Munoz-Price, MD, PhD

Chief Quality and Safety Officer, Virginia Commonwealth University Health System, Richmond, Virginia [162]

Nikhil C. Munshi, MD

Professor of Medicine, Harvard Medical School; Boston VA Healthcare System; Director of Basic and Correlative Sciences; Associate Director, Jerome Lipper Myeloma Center, Dana-Farber Cancer Institute, Boston, Massachusetts [111]

Naoka Murakami, MD, PhD

Instructor in Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [313]

John R. Murphy, PhD

Professor of Medicine, Division of Infectious Diseases, Johns Hopkins School of Medicine, Baltimore, Maryland [150]

Timothy F. Murphy, MD

SUNY Distinguished Professor; Director, UB Clinical and Translational Science Institute; Senior Associate Dean for Clinical and Translational Research, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, The State University of New York, Buffalo, New York [157]

Barbara E. Murray, MD

J. Ralph Meadows Professor of Medicine, Division of Infectious Diseases; Professor, Microbiology and Molecular Genetics, University of Texas Medical School, Houston, Texas [149]

Joseph A. Murray, MD

Professor of Medicine, Departments of Internal Medicine and Immunology, Mayo Clinic School of Medicine, Rochester, Minnesota [46]

Mark B. Mycyk, MD

Associate Professor, Department of Emergency Medicine, Northwestern University Feinberg School of Medicine; Chair of Research, Department of Emergency Medicine, Cook County Health, Chicago, Illinois [459]

Avindra Nath, MD

Chief, Section of Infections of the Nervous System; Clinical Director, National Institute of Neurological Disorders and Stroke (NINDS), National Institutes of Health, Bethesda, Maryland [139]

Edward T. Naureckas, MD

Professor of Medicine, Section of Pulmonary and Critical Care Medicine, University of Chicago, Chicago, Illinois [285]

Eric G. Neilson, MD

Vice President for Medical Affairs; Lewis Landsberg Dean Professor of Medicine and Cell and Molecular Biology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [309, 314, A4]

Tuhina Neogi, MD, PhD

Professor of Medicine and Chief of Rheumatology, Section of Rheumatology, Department of Medicine; Professor of Epidemiology, Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts [371]

Eric J. Nestler, MD, PhD

Nash Family Professor, Department of Neuroscience; Director, Friedman Brain Institute; Dean for Academic and Scientific Affairs, Ichan School of Medicine at Mount Sinai, New York, New York [451]

Hartmut P. H. Neumann, MD

Unit for Preventive Medicine, Department of Nephrology and General Medicine, Albert-Ludwigs University of Freiburg, Freiburg, Germany [387]

Kathleen M. Neuzil, MD, MPH

Director, Center for Vaccine Development & Global Health, University of Maryland School of Medicine, Baltimore, Maryland [200]

Jonathan Newmark, MD, MM

Colonel (Retired), Medical Corps, US Army; Adjunct Professor, Neurology, F. Edward Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland; Clinical Assistant Professor, Neurology, School of Medicine and Health Sciences, George Washington University, Washington, DC; Department of Neurology, Washington DC Veterans' Affairs Medical Center, Washington, DC; Senior Medical Advisor, Office of Biodefense Research and Surety, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland [S4]

J. Curtis Nickel, MD, FRCS(C)

Professor, Department of Urology, Queen's University at Kingston; Staff Urologist, Kingston Health Sciences Centre, Kingston, Ontario, Canada [51]

Michael S. Niederman, MD

Professor of Clinical Medicine, Weill Cornell Medical College; Division of Pulmonary and Critical Care Medicine, New York Presbyterian/Weill Cornell Medical Center, New York, New York [126]

Kevin D. Niswender, MD, PhD

Associate Professor of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee [403]

Scott A. Norton, MD, MPH, MSc

Professor of Dermatology and Pediatrics, George Washington University School of Medicine and Health Sciences; Chief of Dermatology, Division of Dermatology, Children's National Health System, Washington, DC [461]

Emily Nosova, MD

Assistant Professor of Medicine, Division of Endocrinology, Diabetes and Bone Disease, Icahn School of Medicine and Mount Sinai Health System, New York, New York [398]

Thomas B. Nutman, MD

Head, Helminth Immunology Section; Head, Clinical Parasitology Section; Chief, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [232, 233]

Katherine L. O'Brien, MD, MPH

Director, IVB, World Health Organization, Geneva, Switzerland [146]

Max R. O'Donnell, MD, MPH

Associate Professor of Medicine & Epidemiology, Division of Pulmonary, Allergy, and Critical Care Medicine & Department of Epidemiology, Columbia University Irving Medical Center, New York, New York [181]

Nigel O'Farrell, MD, FRCP

Pasteur Suite Ealing Hospital, London, United Kingdom [173]

Jennifer Ogar, MS, CCC-SLP

Speech-Language Pathologist, Memory and Aging Center, University of California, San Francisco, San Francisco, California [V2]

Patrick T. O'Gara, MD

Professor of Medicine, Harvard Medical School; Watkins Family Distinguished Chair in Cardiology, Brigham and Women's Hospital, Boston, Massachusetts [42, 239, 261–268]

C. Warren Olanow, MD, FRCPC, FRCP(hon)

Professor and Chairman Emeritus, Department of Neurology; Professor Emeritus, Department of Neuroscience, Mount Sinai School of Medicine, New York, New York; CEO, Clintrex, LLC [435, 436]

Stephen O'Rahilly, MD, FRS, FMedSci

Professor of Clinical Biochemistry and Medicine and Director of the MRC Metabolic Disease Unit, University of Cambridge, Addenbrookes Hospital, Cambridge, United Kingdom [401]

Joseph G. Ouslander, MD

Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, Florida [477]

Chung Owyang, MD

H. Marvin Pollard Professor of Internal Medicine; Professor of Molecular and Integrative Physiology; Chief, Division of Gastroenterology and Hepatology; Director, Pollard Institute for Medical Research; University of Michigan Health System, Ann Arbor, Michigan [321, 327]

Umesh D. Parashar, MBBS, MPH

Chief, Viral Gastroenteritis Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [203]

Shreyaskumar R. Patel, MD

Robert R. Herring Distinguished Professor of Medicine; Center Medical Director, Sarcoma Center, The University of Texas MD Anderson Cancer Center, Houston, Texas [91]

Gustav Paumgartner, MD

Professor Emeritus of Medicine, University of Munich, Munich, Germany [346]

David A. Pegues, MD

Professor of Medicine, Division of Infectious Diseases, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [165]

Steven A. Pergam, MD, MPH

Associate Professor, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center; Associate Professor, Department of Medicine, Division of Allergy & Infectious Diseases, University of Washington; Medical Director, Infection Prevention, Seattle Cancer Care Alliance, Seattle, Washington [159]

Karran A. Phillips, MD, MSc

Clinical Director, National Institute on Drug Abuse, National Institutes of Health, Baltimore, Maryland [457]

Richard J. Pollack, PhD

Senior Environmental Public Health Officer, Department of Environmental Health and Safety, Harvard University, Cambridge, Massachusetts [461]

Martin R. Pollak, MD

George C. Reisman Professor of Medicine, Harvard Medical School; Beth Israel Deaconess Medical Center, Boston, Massachusetts [315]

Sir Andrew J. Pollard, BSc, MA, MBBS, MRCP(UK), FRCPCH, PhD, DIC, FHEA, FIDSA, FMedSci

Professor of Paediatric Infection and Immunity, Department of Paediatrics, University of Oxford; Children's Hospital, Oxford, United Kingdom [155]

Nongnooch Poowanawittayakom, MD, MPH

Assistant Professor, Division of Infectious Diseases, Allergy, and Immunology, St. Louis University, St. Louis, Missouri [130]

Reuven Porat, MD

Professor of Medicine, Department of Internal Medicine, Tel Aviv Sourasky Medical Center; Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel [18]

John T. Potts, Jr., MD

Jackson Distinguished Professor of Clinical Medicine, Harvard Medical School; Director of Research and Physician-in-Chief Emeritus, Massachusetts General Hospital, Boston, Massachusetts [410]

Lawrie W. Powell, AC, MD, PhD

Professor Emeritus, The University of Queensland and the Royal Brisbane and Women's Hospital, Queensland, Australia [414]

Alvin C. Powers, MD

Joe C. Davis Chair in Biomedical Science; Professor of Medicine, Molecular Physiology and Biophysics; Director, Vanderbilt Diabetes Center; Chief, Division of Diabetes, Endocrinology, and Metabolism, Vanderbilt University Medical Center, Nashville, Tennessee [403–405]

Daniel S. Pratt, MD

Assistant Professor of Medicine, Harvard Medical School; Clinical Director, Liver Transplantation; Director, Autoimmune and Cholestatic Liver Center, Massachusetts General Hospital, Boston, Massachusetts [49, 337, 346]

Michael B. Prentice, MBChB, PhD, FRCP(UK), FRCPath, FFPRCPI

Professor of Medical Microbiology, School of Microbiology, University College Cork, Cork, Ireland [171]

Stanley B. Prusiner, MD

Director, Institute for Neurodegenerative Diseases; Professor, Department of Neurology, UCSF Weill Institute for Neurosciences, University of California, San Francisco; Professor, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California [424, 438]

Thomas C. Quinn, MD, MSc

Professor of Medicine and Pathology, Johns Hopkins University School of Medicine; Director, Johns Hopkins Center for Global Health, Baltimore, Maryland [189]

Gil D. Rabinovici, MD

Ed Fein and Pearl Landrith Distinguished Professor, Memory and Aging Center, Department of Neurology, Department of Radiology and Biomedical Imaging, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California [29, 431, V2]

Daniel J. Rader, MD

Seymour Gray Professor of Molecular Medicine; Chair, Department of Genetics; Chief, Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania [407]

Kanwal Raghav, MBBS, MD

Associate Professor, GI Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas [92]

Kaitlin Rainwater-Lovett, PhD, MPH

Senior Staff Scientist, Asymmetric Operations Sector, Johns Hopkins Applied Physics Laboratory, Laurel, Maryland [205]

Sanjay Ram, MBBS

Professor of Medicine, Division of Infectious Diseases & Immunology, University of Massachusetts Chan Medical School, Worcester, Massachusetts [156]

Prashanth S. Ramachandran, MBBS

Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, California [S2]

Reuben Ramphal, MD

Courtesy Professor, Department of Medicine, University of Florida College of Medicine, Gainesville, Florida [164]

Kathryn Moynihan Ramsey, PhD

Research Assistant Professor, Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [485]

Vikram R. Rao, MD, PhD

Distinguished Professor in Neurology; Associate Professor of Clinical Neurology; Chief, Epilepsy Division, Department of Neurology, University of California, San Francisco, San Francisco, California [425]

Didier Raoult, MD, PhD

Emeritus Professor, IHU Méditerranée Infection, Marseille, France. Aix-Marseille Université, Marseille, France [170]

Kumanan Rasanathan, MBChB, MPH, FAFPHM

Unit Head, Equity and Health (EQH), Department of Social Determinants of Health (SDH), World Health Organization, Phnom Penh, Cambodia [474]

James P. Rathmell, MD

Leroy D. Vandam Professor of Anaesthesia, Harvard Medical School; Chair, Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, Massachusetts [13]

Mario C. Raviglione, MD, FRCP (UK), FERS, Hon RSP (RF)

Full Professor of Global Health; Co-Director, Centre for Multidisciplinary Research in Health Science (MACH), University of Milan, Milan, Italy [178]

Divya Reddy, MBBS, MPH

Associate Professor of Medicine; Program Director, Pulmonary and Critical Care Fellowship; Medical Director, Bronchiectasis and Nontuberculous Mycobacterial (NTM) Disease Program, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York [181]

Susan Redline, MD, MPH

Peter C. Farrell Professor of Sleep Medicine, Harvard Medical School; Brigham and Women's Hospital; Boston, Massachusetts [297]

Sharon L. Reed, MD, MScCTM

Professor of Pathology and Medicine, University of California, San Diego School of Medicine, La Jolla, California [221, 223, S12]

Susan E. Reef, MD

Medical Epidemiologist, Centers for Disease Control and Prevention, Atlanta, Georgia [206]

Michael Regner, MD

Neuroradiology Clinical Instructor, Department of Radiology & Biomedical Imaging, University of California, San Francisco, San Francisco, California [A16]

Victor I. Reus, MD

Distinguished Professor Emeritus, Department of Psychiatry and Behavioral Sciences, University of California, San Francisco School of Medicine; UCSF Weill Institute for Neurosciences, San Francisco, California [452]

Bernardo Reyes, MD

Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, Florida [477]

Joseph J. Rhatigan, MD

Associate Chief, Division of Global Health Equity, Brigham and Women's Hospital; Associate Professor, Harvard Medical School and Harvard T.H. Chan School of Public Health, Boston, Massachusetts [472]

Peter A. Rice, MD

Professor of Medicine, Division of Infectious Diseases & Immunology, University of Massachusetts Chan Medical School, Worcester, Massachusetts [156]

Eugene T. Richardson, MD, PhD

Assistant Professor of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts [475]

Jan H. Richardus, MD, PhD

Professor of Infectious Diseases and Public Health, Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands [179]

Michael R. Rickels, MD, MS

Willard and Rhoda Ware Professor in Diabetes and Metabolic Diseases, Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania [404, 405]

Elizabeth Robbins, MD

Clinical Professor, Pediatrics, Emeritus, University of California, San Francisco, San Francisco, California [S9]

Gary L. Robertson, MD

Emeritus Professor of Medicine, Northwestern University School of Medicine, Chicago, Illinois [381]

Dan M. Roden, MD

Professor of Medicine, Pharmacology, and Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee [67, 68]

James A. Romano, Jr., PhD, DABT, ATS

Principal Senior Life Scientist Advisor, Tunnell Government Services, Inc., Rockville, Maryland [S4]

Karen L. Roos, MD

John and Nancy Nelson Professor of Neurology; Professor of Neurological Surgery, Indiana University School of Medicine, Indianapolis, Indiana [137, 138, 140]

Allan H. Ropper, MD, FRCP, FACP

Professor of Neurology, Harvard Medical School; Deputy Editor, *New England Journal of Medicine*, Boston, Massachusetts [28]

Rossana Rosa, MD

Infectious Diseases Consultant, UnityPoint Clinic, Des Moines, Iowa [162]

Ivan O. Rosas, MD

Professor and Section Chief, Pulmonary, Critical Care, and Sleep Medicine, Baylor College of Medicine, Houston, Texas [293]

Mark Roschewski, MD

Clinical Director, Lymphoid Malignancies Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland [95]

Misha Rosenbach, MD

Associate Professor, Perelman School of Medicine at the University of Pennsylvania, Departments of Dermatology and Internal Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania [60]

Roger N. Rosenberg, MD

Zale Distinguished Chair and Professor of Neurology, Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas [439]

Myrna R. Rosenfeld, MD, PhD

Institut d'Investigacions Biomèdiques August Pi i Sunyer, Fundació Clínic per a la Recerca Biomèdica, Spain; Adjunct Professor, University of Pennsylvania, Philadelphia, Pennsylvania [94]

Deborah C. Rubin, MD

Professor of Medicine and of Developmental Biology; Associate Chair for Faculty Affairs and Director, Womens' GI Committee, Washington University School of Medicine, St. Louis, Missouri [325]

Thomas A. Russo, MD, CM

SUNY Distinguished Professor of Medicine and Microbiology & Immunology, Chief, Division of Infectious Diseases, Jacobs School of Medicine and the Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, New York [161, 175, 176]

George W. Rutherford, MD

Professor of Epidemiology, Preventive Medicine, Pediatrics and History, and Head, Division of Infectious Disease and Global Epidemiology, Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California [473]

Edward T. Ryan, MD

Professor of Medicine, Harvard Medical School; Professor of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health; Director, Global Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts [168]

Manish Sadarangani, MA, BM, BCh, DPhil

Associate Professor, Department of Pediatrics, University of British Columbia; Director, Vaccine Evaluation Center, BC Children's Hospital Research Institute, Vancouver, British Columbia, Canada [155]

David J. Salant, MD

Professor of Medicine, Boston University School of Medicine; Chief, Renal Section, Boston University Medical Center, Boston, Massachusetts [316]

Richard B. Saltman, PhD

Professor of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, Georgia [7]

Blossom Samuels, MD

Attending, Westchester Medical Center; Clinical Assistant Professor, New York Medical College, Valhalla, New York [411]

Martin A. Samuels, MD, DSc (hon), FACP, FAAN, FRCP, FANA

Miriam Sydney Joseph Distinguished Professor of Neurology, Harvard Medical School; Founding Chair Emeritus, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts [V7]

Vaishali Sanchorawala, MD

Professor of Medicine; Director, Amyloidosis Center; Director, Autologous Stem Cell Transplantation Program, Boston University School of Medicine and Boston Medical Center, Boston, Massachusetts [112]

Philippe J. Sansonetti, MD

Professor, Collège de France; Emeritus Professor, Institut Pasteur, Paris, France [166]

Clifford B. Saper, MD, PhD

James Jackson Putnam Professor of Neurology and Neuroscience, Harvard Medical School; Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts [31]

William H. Sauer, MD

Associate Professor of Medicine, Harvard Medical School; Section Chief, Cardiac Arrhythmia Service, Brigham and Women's Hospital, Boston, Massachusetts [243–256, 306]

Edward A. Sausville, MD, PhD

National Cancer Institute, Bethesda, Maryland (Retired); Marlene & Stewart Greenebaum Comprehensive Cancer Center, University of Maryland, Baltimore, Maryland [73]

David T. Scadden, MD

Gerald and Darlene Jordan Professor of Medicine; Chair Emeritus and Professor, Department of Stem Cell and Regenerative Biology, Harvard University; Director, Center for Regenerative Medicine, Massachusetts General Hospital, Co-director, Harvard Stem Cell Institute, Cambridge, Massachusetts [96]

Thomas E. Scammell, MD

Professor, Harvard Medical School; Beth Israel Deaconess Medical Center; Boston Children's Hospital, Boston, Massachusetts [31]

Anthony H. V. Schapira, MD, DSc, FRCP, FMedSci

Head and Professor, Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology; Director of UCL Royal Free Campus; Vice-Dean UCL, London, United Kingdom [435]

Howard I. Scher, MD, FASCO

Professor of Medicine, Weill Cornell Medicine College; D. Wayne Calloway Chair in Urologic Oncology; Head, Biomarker Development Program, Office of the Physician in Chief; Member and Attending Physician, Genitourinary Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York [87]

Gordon Schiff, MD

Associate Professor of Medicine, Harvard Medical School; Associate Director, Brigham and Women's Hospital Center Patient Safety Research; Quality and Safety Director, HMS Center for Primary Care, Boston, Massachusetts [9]

Scott Schissel, MD, PhD

Chief of Medicine, Department of Medicine, Brigham and Women's Faulkner Hospital, Jamaica Plain, Massachusetts [302]

Bernd Schnabl, MD

Professor of Medicine, Department of Medicine, Division of Gastroenterology, University of California San Diego, La Jolla, California [342]

Marc A. Schuckit, MD

Distinguished Professor of Psychiatry, University of California, San Diego Medical School, La Jolla, California [453]

William W. Seeley, MD

Professor of Neurology and Pathology, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California [29, 431–434]

Florencia Pereyra Segal, MD

Assistant Professor of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [141]

Julian L. Seifter, MD

Associate Professor of Medicine, Harvard Medical School; Distinguished Nephrologist, Brigham and Women's Hospital, Boston, Massachusetts [308, 319]

Jaime Sepúlveda, MD, MPH, MSc, DrSc

Haile T. Debas Distinguished Professor of Global Health; Executive Director, Institute for Global Health Sciences, University of California, San Francisco, San Francisco, California [473]

Christopher W. Seymour, MD, MSc

Associate Professor of Critical Care and Emergency Medicine, Department of Critical Care and Emergency Medicine, The CRISMA Center, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania [304]

Majid Shafiq, MD, MPH

Medical Director, Interventional Pulmonology, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [286]

Ankoor Shah, MD

Associate Professor, Department of Medicine, Division of Rheumatology and Immunology, Duke University Medical Center, Durham, North Carolina [358]

Erica S. Shenoy, MD, PhD

Associate Professor of Medicine, Harvard Medical School; Associate Chief, Infection Control Unit, Massachusetts General Hospital, Boston, Massachusetts [144]

Kanade Shinkai, MD, PhD

Professor, Department of Dermatology, University of California, San Francisco, San Francisco, California [60]

Edwin K. Silverman, MD, PhD

Professor of Medicine, Harvard Medical School; Chief, Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [292]

Shakti Singh, MSc, PhD

Research Scientist, The Lundquist Institute at Harbor-UCLA Medical Center, Torrance, California [216]

Karl Skorecki, MD, FCRPC, FASN

Dean, Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel [311, 468]

Wade S. Smith, MD, PhD

Professor of Neurology, Department of Neurology, University of California, San Francisco, San Francisco, California [307, 426–429]

Jeremy Sobel, MD, MPH

Associate Director for Epidemiologic Science, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [153]

Kelly A. Soderberg, PhD

Chief of Staff, Duke Human Vaccine Institute, Department of Medicine, Duke University School of Medicine; Duke University Medical Center, Durham, North Carolina [349, 350]

Scott D. Solomon, MD

The Edward D. Frohlich Distinguished Chair; Professor of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital, Boston, Massachusetts [241, A9]

Julian Solway, MD

Walter L. Palmer Distinguished Service Professor of Medicine and Pediatrics; Dean for Translational Medicine, Biological Sciences Division; Vice Chair for Research, Department of Medicine; Chair, Committee on Molecular Medicine, University of Chicago, Chicago, Illinois [285, 296]

Eric J. Sorscher, MD

Hertz Endowed Professorship, Emory University School of Medicine, Children's Healthcare of Atlanta, Atlanta, Georgia [291]

Brad Spellberg, MD

Professor of Clinical Medicine, Division of Infectious Diseases, Keck School of Medicine at the University of Southern California; Chief Medical Officer, Los Angeles County + University of Southern California (LAC + USC) Medical Center, Los Angeles, California [218]

Jerry L. Spivak, MD

Professor of Medicine and Oncology, Hematology Division, Johns Hopkins University School of Medicine, Baltimore, Maryland [103]

David Spriggs, MD, FACP, FASCO

Faculty Member, Harvard Medical School; Program Director of Gynecologic Oncology, Massachusetts General Hospital Cancer Center, Boston, Massachusetts [89]

E. William St. Clair, MD

W. Lester Brooks, Jr. Professor of Medicine; Professor of Immunology, Department of Medicine, Duke University Medical Center, Durham, North Carolina [358]

John M. Stafford, MD, PhD

Associate Professor of Medicine, Diabetes and Endocrinology, Vanderbilt University School of Medicine; Tennessee Valley Health System, Veterans Affairs, Nashville, Tennessee [405]

Matthew W. State, MD, PhD

Oberndorf Family Distinguished Professor; Chair, Department of Psychiatry and Behavioral Sciences; President, Langley Porter Psychiatric Hospital and Clinics, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California [451]

Allen C. Steere, MD

Professor of Medicine, Harvard Medical School and Massachusetts General Hospital, Boston, Massachusetts [186]

Martin H. Steinberg, MD

Professor of Medicine, Pediatrics, Pathology and Laboratory Medicine, Boston University School of Medicine, Boston, Massachusetts [98]

Dennis L. Stevens, MD, PhD

Professor of Medicine, University of Washington School of Medicine, Seattle, Washington; Director, Center of Biomedical Research Excellence in Emerging/Reemerging Infectious Diseases, Boise Veterans Affairs Medical Center, Boise, Idaho [129, 154]

Lynne Warner Stevenson, MD

Professor of Medicine; Program Director, Advanced Heart Failure Fellowship Program, Vanderbilt University Medical Center, Nashville, Tennessee [259]

Benjamin K. Stoff, MD, MAB

Associate Professor of Dermatology, Emory University School of Medicine; Senior Faculty Fellow, Emory Center for Ethics, Atlanta, Georgia [A5]

John H. Stone, MD, MPH

Professor of Medicine, Harvard Medical School; The Edward Fox Chair in Medicine, Massachusetts General Hospital, Boston, Massachusetts [368]

Shyam Sundar, MD

Distinguished Professor, Department of Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India [226]

Neeraj K. Surana, MD, PhD

Assistant Professor of Pediatrics, Molecular Genetics and Microbiology, and Immunology, Duke University, Durham, North Carolina [18, 119, 177, 471]

Paolo M. Suter, MD, MS

Department of Endocrinology, Diabetology, and Clinical Nutrition, University Hospital, Zurich, Switzerland [333]

Robert A. Swerlick, MD

Department of Dermatology, Emory University School of Medicine, Atlanta, Georgia [57]

Geoffrey Tabin, MD

Director, Department of Ophthalmology, Stanford University, Stanford, California [462]

Maria Carmela Tartaglia, MD

Associate Professor, Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada [V2]

Joel D. Taurog, MD

Professor of Internal Medicine (Retired), Rheumatic Diseases Division, University of Texas Southwestern Medical Center, Dallas, Texas [362]

Usha B. Tedrow, MD, MSc

Associate Professor of Medicine, Harvard Medical School; Director, Clinical Cardiac Electrophysiology Fellowship; Clinical Director, Ventricular Arrhythmia Program, Brigham and Women's Hospital, Boston, Massachusetts [252–256]

Ayalew Tefferi, MD

Barbara Woodward Lips Professor of Medicine and Hematology, Mayo Clinic, Rochester, Minnesota [110]

Stephen C. Textor, MD

Professor of Medicine, Division of Nephrology and Hypertension, Mayo Clinic School of Medicine, Rochester, Minnesota [278]

R. V. Thakker, MD, ScD, FRCP, FRCPath, FRS, FMedSci

May Professor of Medicine, Academic Endocrine Unit, University of Oxford; O.C.D.E.M., Churchill Hospital, Headington, Oxford, United Kingdom [388]

Holger Thiele, MD

Full Professor of Internal Medicine/Cardiology; Director, Heart Center, Department of Internal Medicine/Cardiology, University of Leipzig, Leipzig, Germany [305]

C. Louise Thwaites, MBBS, BSc, MD

Clinical Research Fellow, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; Clinical Lecturer, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United Kingdom [152]

Pierre Tiberghien, MD, PhD

Professor of Medicine, Bourgogne Franche-Comté University, Besançon; Senior Advisor, Etablissement Français du Sang, Paris, France [113]

Zelig A. Tochner, MD

Professor Emeritus of Radiation Oncology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania [S5]

Karina A. Top, MD, MSc

Associate Professor of Pediatrics and Community Health & Epidemiology, Dalhousie University, Halifax, Nova Scotia, Canada [160]

Mark Topazian, MD

Professor of Medicine, Mayo Clinic, Rochester, Minnesota [322, V5]

Camilo Toro, MD

Director, Adult NIH Undiagnosed Diseases Program, National Institutes of Health, Bethesda, Maryland [492]

Barbara W. Trautner, MD, PhD

Professor, Section of Infectious Diseases, Department of Medicine, Baylor College of Medicine; Investigator, Houston VA Center for Innovations in Quality, Effectiveness and Safety (IQuEST), Houston, Texas [135]

Katherine L. Tuttle, MD

Assistant Professor, Department of Pediatrics, Pediatric Allergy/Immunology; Assistant Professor, Department of Medicine, Allergy/Immunology and Rheumatology (SMD), University of Rochester Medical Center, Rochester, New York [352]

Kenneth L. Tyler, MD

Louise Baum Endowed Chair and Chairman of Neurology; Professor of Medicine and Immunology-Microbiology, University of Colorado School of Medicine, Aurora, Colorado; Neurologist, Rocky Mountain VA Medical Center, Aurora, Colorado [137, 138, 140]

Elizabeth R. Unger, PhD, MD

Division of High-Consequence Pathogens and Pathology, National Center for Zoonotic and Emerging Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [450]

Prashant Vaishnav, MD

Assistant Professor of Medicine; Director of Quality Assurance and Inpatient Services, Mount Sinai Heart, Mount Sinai Hospital, Icahn School of Medicine at Mount Sinai, New York, New York [480]

Anne Marie Valente, MD

Associate Professor of Medicine and Pediatrics, Harvard Medical School; Director, Boston Adult Congenital Heart Disease and Pulmonary Hypertension Program, Boston Children's Hospital, Brigham and Women's Hospital, Boston, Massachusetts [269]

Wim H. van Brakel, MD, MSc, PhD

Medical Director, NLR, Amsterdam, The Netherlands [179]

Jos W. M. van der Meer, MD, PhD

Emeritus Professor of Medicine, Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands [20]

Mathew G. Vander Heiden, MD, PhD

Professor and Director, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts [489]

Edouard Vannier, PharmD, PhD

Assistant Professor of Medicine, Division of Geographic Medicine and Infectious Diseases, Department of Medicine, Tufts Medical Center and Tufts University School of Medicine, Boston, Massachusetts [225]

Gauri R. Varadhachary, MD

Professor, GI Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas [92]

John Varga, MD

Frederick Huetwell Professor; Chief, Division of Rheumatology, University of Michigan, Ann Arbor, Michigan [360]

David J. Vaughn, MD

Genitourinary Medical Oncology Professor, Perelman School of Medicine at the University of Pennsylvania, Perelman Center for Advanced Medicine, Philadelphia, Pennsylvania [88]

Birgitte Jyding Vennervold, MD, MSA

Professor, Section for Parasitology and Aquatic Pathobiology, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg, Denmark [234]

John T. Vetto, MD, FACS

Professor of Surgery, Division of Surgical Oncology; Director, Cutaneous Oncology Program, Department of Surgery, Oregon Health & Science University; Program Leader, Melanoma Disease Site Team, OHSU Knight Cancer Institute, Portland, Oregon [76]

Luciano Villarinho, MD

Neuroradiologist, South County Hospital, Wakefield, Rhode Island [A16]

Bert Vogelstein, MD

Professor, Ludwig Center for Cancer Genetics and Therapeutics, Johns Hopkins University School of Medicine; Investigator, Howard Hughes Medical Institute, Baltimore, Maryland [71]

Everett E. Vokes, MD

John E. Ultmann Professor; Chairman, Department of Medicine; Physician-in-Chief, University of Chicago Medicine and Biological Sciences, Chicago, Illinois [77]

Tamara J. Vokes, MD

Professor, Department of Medicine, Section of Endocrinology, University of Chicago, Chicago, Illinois [412]

Nora D. Volkow, MD

Director, National Institute on Drug Abuse (NIDA), National Institutes of Health, Rockville, Maryland [455]

Kevin G. Volpp, MD, PhD

Director, Penn Center for Health Incentives and Behavioral Economics; Founders Presidential Distinguished Professor, Perelman School of Medicine and the Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania [481]

Daniel D. Von Hoff, MD, FACP, FASCO, FAACR

Distinguished Professor, Translational Genomics Research Institute (TGEN), Phoenix, Arizona; Virginia G. Piper Distinguished Chair for Innovative Cancer Research and Chief Scientific Officer, Honor Health Research Institute; Senior Consultant-Clinical Investigations, City of Hope; Professor of Medicine, Mayo Clinic, Scottsdale, Arizona [83]

Martin H. Voss, MD

Clinical Director, Genitourinary Oncology Service, Memorial Sloan Kettering Cancer Center, New York, New York [85]

Jiri F. P. Wagenaar, MD, PhD

Internist and Infectious Disease Specialist, Northwest Clinics, Alkmaar, The Netherlands [184]

Jesse Waggoner, MD

Assistant Professor, Department of Medicine, Division of Infectious Diseases, Emory University, Atlanta, Georgia [124]

Sushrut S. Waikar, MD, MPH

Chief, Section of Nephrology; Norman G. Lewinsky Professor of Medicine, Boston University School of Medicine, Boston, Massachusetts [310]

Matthew K. Waldor, MD, PhD

Edward H. Kass Professor of Medicine, Harvard Medical School, Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts [168]

David H. Walker, MD

The Carmage and Martha Walls Distinguished University Chair in Tropical Diseases; Professor, Department of Pathology; Executive Director, Center for Biodefense and Emerging Infectious Diseases, University of Texas Medical Branch, Galveston, Texas [187]

Mark F. Walker, MD

Associate Professor, Neurology, Case Western Reserve University; Director, Daroff-Dell'Osso Ocular Motility Laboratory, VA Northeast Ohio Healthcare System, Cleveland, Ohio [22]

George R. Washko, MD, MMSc

Associate Professor of Medicine, Harvard Medical School; Associate Physician, Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [286, A12]

Michael E. Wechsler, MD, MMSc

Professor of Medicine; Director, Asthma Program, Department of Medicine, National Jewish Health, Denver, Colorado [288]

Anthony P. Weetman, MD, DSc

University of Sheffield, School of Medicine, Sheffield, United Kingdom [382–385]

Robert A. Weinstein, MD

The C. Anderson Hedberg MD Professor of Internal Medicine, Rush University Medical Center; Chairman of Medicine, Emeritus, Cook County Health, Chicago, Illinois [142]

Jeffrey I. Weitz, MD, FRCP(C), FRSC, FACP

Professor of Medicine and Biochemistry and Biomedical Sciences, McMaster University; Executive Director, Thrombosis and Atherosclerosis Research Institute, Hamilton, Ontario, Canada [118]

Peter F. Weller, MD

William Bosworth Castle Professor of Medicine, Harvard Medical School; Professor of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health; Chief Emeritus, Infectious Diseases Division and Vice Chair of Research, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts [229–233, 235]

Andrew Wellman, MD, PhD

Associate Professor of Medicine, Harvard Medical School; Director, Sleep Disordered Breathing Lab, Brigham and Women's Hospital, Boston, Massachusetts [297]

Patrick Y. Wen, MD

Professor of Neurology, Harvard Medical School; Director, Center for Neuro-Oncology, Dana-Farber Cancer Institute; Director, Division of Neuro-Oncology, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts [90]

Michael R. Wessels, MD

John F. Enders Professor of Pediatrics and Professor of Medicine, Harvard Medical School; Senior Physician, Division of Infectious Diseases, Boston Children's Hospital, Boston, Massachusetts [148]

L. Joseph Wheat, MD

Medical Director, MiraVista Diagnostics, Indianapolis, Indiana [212]

A. Clinton White, Jr., MD, FACP, FIDSA, FASTMH

Professor, Infectious Disease Division, Department of Internal Medicine, University of Texas Medical Branch, Galveston, Texas [235]

Nicholas J. White, DSc, MD, FRCP, F Med Sci, FRS

Professor of Tropical Medicine, Mahidol and Oxford Universities, Bangkok, Thailand [224, A2]

Richard J. Whitley, MD

Loeb Eminent Scholar in Pediatrics; Professor of Pediatrics, Microbiology and Neurosurgery, The University of Alabama at Birmingham, Birmingham, Alabama [193]

xxxviii Eleanor Wilson, MD, MHS

Associate Professor of Medicine, Associate Director of Clinical Research, Division of Clinical Care and Research, Institute of Human Virology, University of Maryland School of Medicine, Baltimore, Maryland [191]

Michael R. Wilson, MD, MAS

Rachleff Family Distinguished Associate Professor in Neurology, University of California San Francisco Weill Institute for Neurosciences; Staff Physician, University of California San Francisco Medical Center and Zuckerberg San Francisco General Hospital, San Francisco, California [137, 139, S2]

Bruce U. Wintroub, MD

Professor and Chair, Department of Dermatology, University of California, San Francisco, San Francisco, California [60]

Allan W. Wolfson, MD

The Herman Lopata Chair in Liver Disease Research; Professor of Medicine and Anatomy and Structural Biology; Associate Chair of Medicine for Research; Chief, Division of Hepatology; Director, Marion Bessin Liver Research Center, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York [338]

Louis Michel Wong Kee Song, MD

Professor of Medicine, Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, Minnesota [322, V5]

John B. Wong, MD

Professor of Medicine, Tufts University School of Medicine; Interim Chief Scientific Officer, Tufts Medical Center, Boston, Massachusetts [4]

Thomas E. Wood, PhD

Research Fellow, Department of Medicine, Division of Infectious Diseases, Massachusetts General Hospital; Department of Microbiology, Harvard Medical School, Boston, Massachusetts [120]

Jennifer A. Woyach, MD

Professor of Medicine, Division of Hematology, The Ohio State University, Columbus, Ohio [107]

Peter F. Wright, MD

Professor of Pediatrics, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire [200]

Henry M. Wu, MD, DTM&H, FIDSA

Associate Professor of Medicine, Division of Infectious Diseases, Emory University; Director, Emory TravelWell Center, Atlanta, Georgia [124]

Kim B. Yancey, MD

Professor and Chair, Department of Dermatology, University of Texas Southwestern Medical Center in Dallas, Dallas, Texas [56, 59]

Lonny Yarmus, DO, MBA

Professor of Medicine, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland [299]

Yusuf Yazici, MD

Clinical Associate Professor of Medicine, New York University Grossman School of Medicine, New York, New York [364]

Baligh R. Yehia, MD, MPP, MSc

Ascension Health, St. Louis, Missouri [400]

Janet A. Yellowitz, DMD, MPH

Associate Professor; Director, Special Care and Geriatric Dentistry, University of Maryland School of Dentistry, Baltimore, Maryland [A3]

Lam Minh Yen, MD

Senior Clinical Researcher, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam [152]

Neal S. Young, MD

Chief, Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland [102, 469]

Paul C. Zei, MD, PhD

Associate Professor of Medicine, Harvard Medical School; Director, Clinical Atrial Fibrillation Program, Brigham and Women's Hospital, Boston, Massachusetts [243, 246–251]

Jing Zhou, MD, PhD, FASN

Professor of Medicine, Harvard Medical School; Director, Laboratory of Molecular Genetics and Developmental Biology of Disease, Renal Division; Director, Center for Polycystic Kidney Disease Research, Brigham and Women's Hospital; Boston, Massachusetts [315]

Werner Zimmerli, MD

Professor of Medicine, Basel University, Interdisciplinary Unit of Orthopaedic Infections, Kantonsspital Baselland, Liestal, Switzerland [131]

Laura A. Zimmerman, MPH

Epidemiologist, Centers for Disease Control and Prevention, Atlanta, Georgia [206]

The Editors are pleased to present the 21st edition of *Harrison's Principles of Internal Medicine*. This 21st edition is a true landmark in medicine, spanning 71 years and multiple generations of trainees and practicing clinicians. While medicine and medical education have evolved, readers will appreciate how this classic textbook has retained enduring features that have distinguished it among medical texts—a sharp focus on the clinical presentation of disease, expert in-depth summaries of pathophysiology and treatment, and highlights of emerging frontiers of science and medicine. Indeed, *Harrison's* retains its conviction that, in the profession of medicine, we are all perpetual students with lifelong learning as our common goal.

Harrison's is intended for learners throughout their careers. For *students*, Part 1, Chapter 1 begins with an overview of "The Practice of Medicine." In this introductory chapter, the editors continue the tradition of orienting clinicians to the *science* and the *art* of medicine, emphasizing the values of our profession while incorporating new advances in technology, science, and clinical care. Part 2, "Cardinal Manifestations and Presentation of Diseases," is a signature feature of *Harrison's*. These chapters eloquently describe how patients present with common clinical conditions, such as headache, fever, cough, palpitations, or anemia, and provide an overview of typical symptoms, physical findings, and differential diagnosis. Mastery of these topics prepares students for subsequent chapters on specific diseases they will encounter in courses on pathophysiology and in clinical clerkships. For *residents* and *fellows* caring for patients and preparing for board examinations, *Harrison's* remains a definitive source of trusted content written by internationally renowned experts. Trainees will be reassured by the depth of content, comprehensive tables, and illuminating figures and clinical algorithms. Many examination questions are based on key testing points derived from *Harrison's* chapters. A useful companion book, *Harrison's Self-Assessment and Board Review*, includes over 1000 questions, offers comprehensive explanations of the correct answer, and provides links to the relevant chapters in the textbook. *Practicing clinicians* must keep up with an ever-changing knowledge base and clinical guidelines as part of lifelong learning. Clinicians can trust that chapters are updated extensively with each edition of *Harrison's*. The text is an excellent point-of-care reference for clinical questions, differential diagnosis, and patient management. In addition to the expanded and detailed Treatment sections, *Harrison's* continues its tradition of including "Approach to the Patient" sections, which provide an expert's overview of the practical management of common but often complex clinical conditions.

This edition has been modified extensively in its structure as well as its content and offers a more consistently standardized format for each disease chapter. The authors and editors have curated rigorously and synthesized the vast amount of information that comprises general internal medicine—and each of the major specialties—into a highly readable and informative two-volume book. Readers will appreciate the concise writing style and substantive quality that have always characterized *Harrison's*. This book has a sharp focus on essential information with a goal of providing clear and definitive answers to clinical questions.

In the 21st edition, examples of new chapters include "Precision Medicine and Clinical Care," focusing on the ever-growing pool of "big data" used to provide individualized genotype-phenotype correlations; "Mechanisms of Regulation and Dysregulation of the Immune System," focusing on the extraordinary advances made over the past 5 years in understanding the complex and subtle mechanisms whereby the immune system is regulated and how perturbations in this regulation lead to disease states as well as targets for therapeutic intervention; new chapters on Alzheimer's disease and related conditions, with a special focus on vascular dementia, a common and treatable cause of cognitive

loss; and a new chapter on marijuana and marijuana use disorders, as well as updated management guidelines for multiple sclerosis and the expanding array of other autoimmune nervous system diseases that can now be identified and treated.

Other new chapters include "Vaccine Opposition and Hesitancy," "Precision Medicine and Clinical Care," "Diagnosis: Reducing Errors and Improving Quality," "Approach to the Patient with Renal or Urinary Tract Disease," "Interventional Nephrology," "Health Effects of Climate Change," and "Circulating Nucleic Acids as Liquid Biopsies and Noninvasive Disease Biomarkers." In addition, many chapters have new authors.

The chapter, "Vaccine Opposition and Hesitancy," provides an overview of the current antivaccination crisis, the issues involved, and specific strategies to utilize within the clinical setting to address the lack of confidence that many patients feel toward the health care system. The chapter, "Metabolomics," outlines an emerging and important new and sensitive approach to measuring perturbations within a system or patient that will likely become a routine part of the clinical armamentarium for diagnosing, monitoring, and treating disease.

In addition to these and other new topics, the 21st edition presents important updates in the established chapters, such as the microbiology and clinical management of SARS-CoV-2 infection, the use of gene editing for sickle cell anemia and thalassemia, gene therapy for hemophilia, new immunotherapies for autoimmune diseases and cancers, and novel approaches to vaccine development, among many others. Our focus on forward-looking issues of emerging clinical importance continues with the series of chapters entitled "Frontiers," which foreshadows cutting-edge science that will change medical practice in the near term. Examples of new Frontier chapters include "Machine Learning and Augmented Intelligence," "Metabolomics," "Protein Folding Disorders," and "Novel Approaches to Disease of Unknown Etiology."

Harrison's content is available in a variety of print and digital formats, including eBooks, apps, and a popular, widely used online platform available at www.accessmedicine.com.

We have many people to thank for their efforts in producing this book. First, the authors have done a superb job of producing authoritative chapters that synthesize vast amounts of scientific and clinical data to create informative and practical approaches to managing patients. In today's information-rich, rapidly evolving environment, they have ensured that this information is current. We are most grateful to our colleagues who work closely with each editor to facilitate communication with the authors and help us keep *Harrison's* content current. In particular, we wish to acknowledge the expert support of Lauren Bauer, Patricia Conrad, Patricia L. Duffey, Gregory K. Folkers, Julie B. McCoy, Elizabeth Robbins, Marie Scurti, and Stephanie C. Tribuna. Scott Grillo and James Shanahan, our long-standing partners at McGraw Hill's Professional Publishing group, have inspired the creative and dynamic evolution of *Harrison's*, guiding the development of the book and its related products in new formats. Kim Davis, as Managing Editor, has adeptly ensured that the complex production of this multi-authored textbook proceeded smoothly and efficiently. Priscilla Beer oversaw the production of our videos and animations; Jeffrey Herzlich, Elleanore Waka, and Rachel Norton, along with other members of the McGraw Hill staff; and Revathi Viswanathan of KnowledgeWorks Global Ltd., shepherded the production of this new edition.

We are privileged to have compiled this 21st edition and are enthusiastic about all that it offers our readers. We learned much in the process of editing *Harrison's* and hope that you will find this edition uniquely valuable as a clinical and educational resource.

The Editors

Related Harrison's Resources

A complete collection to meet your educational, clinical, and board prep needs.

Harrison's Online

The online edition of *Harrison's* is available at www.accessmedicine.com. It requires an institutional or individual subscription separate from the purchase of the print book. The online edition of *Harrison's* features all the chapters from the print edition, plus more than two dozen supplementary chapters in print, atlas, and video formats. *Harrison's Online* includes numerous monthly updates, from the editors of *Harrison's*, on important new developments in medical research and practice. Easily search across the entire *Harrison's* content set, download images and tables for presentations and lectures, view step-by-step videos on common clinical procedures, access the text of the *Harrison's Manual of Medicine*, set up a personalized test exam for board prep, get access to chapters from new editions of *Harrison's* months before book publication, and more.

The Harrison's Manual of Medicine

The *Harrison's Manual of Medicine* provides high-yield, rapid-access clinical summaries of *Harrison's* content, suitable for use at the bedside. Chapters in the *Manual* reflect those likely to be encountered in both the inpatient and outpatient setting. The format is built for ease of use. The *Manual* is available in print, eBook, and app. In addition, the full text of the *Manual* is available to subscribers at accessmedicine.com. This format provides flexibility of format to customers, who can move back and forth between the full scope of *Harrison's Principles of Internal Medicine* and the high-yield clinical essentials of the *Manual*.

The *Manual* includes more than 200 chapters in 17 sections and covers presenting signs and symptoms and major conditions seen in both inpatient and outpatient settings. The full table of contents is available at www.accessmedicine.com.

The Harrison's Self-Assessment and Board Review

This practical resource provides more than 1000 self-assessment questions, most in board-style clinical vignette format with multiple choice answers. The explanations for the questions are comprehensive and provide detailed guidance on correct and incorrect answers. Question-and-answer sets include references to related chapters in *Harrison's Principles of Internal Medicine* for more comprehensive understanding. Use this very handy resource for primary and recertification exam prep, for rotational shelf exams, and for general assessment of understanding of the principles of clinical medicine. This resource is available as a print book, an eBook, an app, and on accessmedicine.com, where users can create personalized testing experiences and receive instant scores on practice tests.

Harrison's Podclass

Our podcast presents bi-weekly episodes covering clinical vignettes across internal medicine, with two expert discussants reviewing common and challenging patient presentations and a series of self-assessment Q&A choices tied to each case. The hosts work through correct and incorrect answer choices and summarize cases with practical pearls that all students and clinicians will find helpful and interesting. *Harrison's Podclass* is available in most of the common podcast outlets and on www.accessmedicine.com.

1

The Practice of Medicine

The Editors



ENDURING VALUES OF THE MEDICAL PROFESSION

No greater opportunity, responsibility, or obligation can fall to the lot of a human being than to become a physician. In the care of the suffering, [the physician] needs technical skill, scientific knowledge, and human understanding. Tact, sympathy, and understanding are expected of the physician, for the patient is no mere collection of symptoms, signs, disordered functions, damaged organs, and disturbed emotions. [The patient] is human, fearful, and hopeful, seeking relief, help, and reassurance.

—Harrison's Principles of Internal Medicine, 1950

The practice of medicine has changed in significant ways since the first edition of this book was published in 1950. The advent of molecular genetics, sophisticated new imaging techniques, robotics, and advances in bioinformatics and information technology have contributed to an explosion of scientific information that has changed fundamentally the way physicians define, diagnose, treat, and attempt to prevent disease. This growth of scientific knowledge continues to evolve at an accelerated pace.

The widespread use of electronic medical records and the Internet have altered the way physicians and other health care providers access and exchange information as a routine part of medical education and practice (Fig. 1-1). As today's physicians strive to integrate an ever-expanding body of scientific knowledge into everyday practice, it is critically important to remember two key principles: first, the ultimate goal of medicine is to prevent disease and, when it occurs, to diagnose it early and provide effective treatment; and second, despite 70 years of scientific advances since the first edition of this text, a trusting relationship between physician and patient still lies at the heart of effective patient care.

THE SCIENCE AND ART OF MEDICINE

Deductive reasoning and applied technology form the foundation for the approach and solution to many clinical problems. Extraordinary advances in biochemistry, cell biology, immunology, and genomics,



FIGURE 1-1 *The Doctor* by Luke Fildes depicts the caring relationship between this Victorian physician and a very ill child. Painted in 1891, the painting reflects the death of the painter's young son from typhoid fever and was intended to reflect the compassionate care provided by the physician even when his tools were not able to influence the course of disease. (Source: History and Art Collection/Alamy Stock Photo.)

coupled with newly developed imaging techniques, provide a window into the most remote recesses of the body and allow access to the innermost parts of the cell. Revelations about the nature of genes and single cells have opened a portal for formulating a new molecular basis for the physiology of systems. Researchers are deciphering the complex mechanisms by which genes are regulated, and increasingly, physicians are learning how subtle changes in many different genes, acting in an integrative contextual way, can affect the function of cells and organisms. Clinicians have developed a new appreciation of the role of stem cells in normal tissue function, in the development of cancer and other disorders, and in the treatment of certain diseases. Entirely new areas of research, including studies of the human microbiome, epigenetics, and noncoding RNAs as regulatory features of the genome, have become important for understanding both health and disease. Information technology enables the interrogation of medical records from millions of individuals, yielding new insights into the etiology, characteristics, prognosis, and stratification of many diseases. With the increasing availability of very large data sets ("big data") from omic analyses and the electronic medical record, there is now a growing need for machine learning and artificial intelligence for unbiased analyses that enhance clinical predictive accuracy. The knowledge gleaned from the *science of medicine* continues to enhance the understanding by physicians of complex pathologic processes and to provide new approaches to disease prevention, diagnosis, and treatment. With continued refinement of unique omic signatures coupled with nuanced clinical pathophenotypes, the profession moves ever closer to practical precision medicine. Yet, skill in the most sophisticated applications of laboratory technology and in the use of the latest therapeutic modality alone does not make a good physician. Extraordinary advances in vaccine platform technology and the use of cryo-electron microscopy for the structure-based design of vaccine immunogens have transformed the field of vaccinology, resulting in the unprecedented speed and success with which COVID-19 vaccines were developed.

When a patient poses challenging clinical problems, an effective physician must be able to identify the crucial elements in a complex history and physical examination; order the appropriate laboratory, imaging, and diagnostic tests; and extract the key results from densely populated computer screens to determine whether to treat or to "watch." As the number of tests increases, so does the likelihood that some incidental finding, completely unrelated to the clinical problem at hand, will be uncovered. Deciding whether a clinical clue is worth pursuing or should be dismissed as a "red herring" and weighing whether a proposed test, preventive measure, or treatment entails a greater risk than the disease itself are essential judgments that a skilled clinician must make many times each day. This combination of medical knowledge, intuition, experience, and judgment defines the *art of medicine*, which is as necessary to the practice of medicine and the precision medicine of the future as is a sound scientific base, and as important for contemporary medical practice as it has been in earlier eras.

CLINICAL SKILLS

History-Taking The recorded history of an illness should include all the facts of medical significance in the life of the patient. Recent events should be given the most attention. Patients should, at some early point, have the opportunity to tell their own story of the illness without frequent interruption and, when appropriate, should receive expressions of interest, encouragement, and empathy from the physician. Any event related by a patient, however trivial or seemingly irrelevant, may provide the key to solving the medical problem. A methodical review of systems is important to elicit features of an underlying disease that might not be mentioned in the patient's narrative. In general, patients who feel comfortable with the physician will offer more complete information; thus, putting the patient at ease contributes substantially to obtaining an adequate history.

An informative history is more than eliciting an orderly listing of symptoms. By listening to patients and noting the ways in which they describe their symptoms, physicians can gain valuable insight. Inflections of voice, facial expression, gestures, and attitude (i.e., “body language”) may offer important clues to patients’ perception of and reaction to their symptoms. Because patients vary considerably in their medical sophistication and ability to recall facts, the reported medical history should be corroborated whenever possible. The social history also can provide important insights into the types of diseases that should be considered and can identify practical considerations for subsequent management. The family history not only identifies rare genetic disorders or common exposures, but often reveals risk factors for common disorders, such as coronary heart disease, hypertension, autoimmunity, and asthma. A thorough family history may require input from multiple relatives to ensure completeness and accuracy. An experienced clinician can usually formulate a relevant differential diagnosis from the history alone, using the physical examination and diagnostic tests to narrow the list or reveal unexpected findings that lead to more focused inquiry.

The very act of eliciting the history provides the physician with an opportunity to establish or enhance a unique bond that can form the basis for a good patient–physician relationship. This process helps the physician develop an appreciation of the patient’s view of the illness, the patient’s expectations of the physician and the health care system, and the financial and social implications of the illness for the patient. Although current health care settings may impose time constraints on patient visits, it is important not to rush the encounter. A hurried approach may lead patients to believe that what they are relating is not of importance to the physician, and, as a result, they may withhold relevant information. The confidentiality of the patient–physician relationship cannot be overemphasized.

Physical Examination The purpose of the physical examination is to identify physical signs of disease. The significance of these objective indications of disease is enhanced when they confirm a functional or structural change already suggested by the patient’s history. At times, however, physical signs may be the only evidence of disease and may not have been suggested by the history.

The physical examination should be methodical and thorough, with consideration given to the patient’s comfort and modesty. Although attention is often directed by the history to the diseased organ or part of the body, the examination of a new patient must extend from head to toe in an objective search for abnormalities. The results of the examination, like the details of the history, should be recorded at the time they are elicited—not hours later, when they are subject to the distortions of memory. Physical examination skills should be learned under direct observation of experienced clinicians. Even highly experienced clinicians can benefit from ongoing coaching and feedback. Simulation laboratories and standardized patients play an increasingly important role in the development of clinical skills. Although the skills of physical diagnosis are acquired with experience, it is not merely technique that determines success in identifying signs of disease. The detection of a few scattered petechiae, a faint diastolic murmur, or a small mass in the abdomen is not a question of keener eyes and ears or more sensitive fingers, but of a mind alert to those findings. Because physical findings can change with time, the physical examination should be repeated as frequently as the clinical situation warrants.

Given the many highly sensitive diagnostic tests now available (particularly imaging techniques), it may be tempting to place less emphasis on the physical examination. Some are critical of physical diagnosis based on perceived low levels of specificity and sensitivity. Indeed, many patients are seen by consultants only after a series of diagnostic tests have been performed and the results are known. This fact should not deter the physician from performing a thorough physical examination since important clinical findings may have escaped detection by diagnostic tests. Especially important, a thorough and thoughtful physical examination may render a laboratory finding unimportant (i.e., certain echocardiographic regurgitant lesions). The act of a hands-on examination of the patient also offers an opportunity

for communication and may have reassuring effects that foster the patient–physician relationship.

Diagnostic Studies Physicians rely increasingly on a wide array of laboratory and imaging tests to make diagnoses and ultimately to solve clinical problems; however, such information does not relieve the physician from the responsibility of carefully observing and examining the patient. It is also essential to appreciate the limitations of diagnostic tests. By virtue of their apparent precision, these tests often gain an aura of certainty regardless of the fallibility of the tests themselves, the instruments used in the tests, and the individuals performing or interpreting the tests. Physicians must weigh the expense involved in laboratory procedures against the value of the information these procedures are likely to provide.

Single laboratory tests are rarely ordered. Instead, physicians generally request “batteries” of multiple tests, which often prove useful and can be performed with a single specimen at relatively low cost. For example, abnormalities of hepatic function may provide the clue to nonspecific symptoms such as generalized weakness and increased fatigability, suggesting a diagnosis of chronic liver disease. Sometimes a single abnormality, such as an elevated serum calcium level, points to a particular disease, such as hyperparathyroidism.

The thoughtful use of screening tests (e.g., measurement of low-density lipoprotein cholesterol) may allow early intervention to prevent disease (*Chap. 6*). Screening tests are most informative when they are directed toward common diseases and when their results indicate whether other potentially useful—but often costly—tests or interventions are needed. On the one hand, biochemical measurements, together with simple laboratory determinations such as routine serum chemistries, blood counts, and urinalysis, often provide a major clue to the presence of a pathologic process. On the other hand, the physician must learn to evaluate occasional screening-test abnormalities that do not necessarily connote significant disease. An in-depth workup after the report of an isolated laboratory abnormality in a person who is otherwise well is often wasteful and unproductive. Because so many tests are performed routinely for screening purposes, it is not unusual for one or two values to be slightly abnormal. Nevertheless, even if there is no reason to suspect an underlying illness, tests yielding abnormal results ordinarily are repeated to rule out laboratory error. If an abnormality is confirmed, it is important to consider its potential significance in the context of the patient’s condition and other test results.

There is almost continual development of technically improved imaging studies with greater sensitivity and specificity. These tests provide remarkably detailed anatomic information that can be pivotal in informing medical decision-making. MRI, CT, ultrasonography, a variety of isotopic scans, and positron emission tomography (PET) have supplanted older, more invasive approaches and opened new diagnostic vistas. In light of their capabilities and the rapidity with which they can lead to a diagnosis, it is tempting to order a battery of imaging studies. All physicians have had experiences in which imaging studies revealed findings that led to an unexpected diagnosis. Nonetheless, patients must endure each of these tests, and the added cost of unnecessary testing is substantial. Furthermore, investigation of an unexpected abnormal finding may lead to an iatrogenic complication or to the diagnosis of an irrelevant or incidental problem. A skilled physician must learn to use these powerful diagnostic tools judiciously, always considering whether the results will alter management and benefit the patient.

■ MANAGEMENT OF PATIENT CARE

Team-Based Care Medical practice has long involved teams, particularly physicians working with nurses and, more recently, with physician assistants and nurse practitioners. Advances in medicine have increased our ability to manage very complex clinical situations (e.g., intensive care units [ICUs], bone marrow transplantation) and have shifted the burden of disease toward chronic illnesses. Because an individual patient may have multiple chronic diseases, he or she may be cared for by several specialists as well as a primary care physician. In the inpatient setting, care may involve multiple consultants along with

the primary admitting physician. Communication through the medical record is necessary but not sufficient, particularly when patients have complex medical problems or when difficult decisions need to be made about the optimal management plan. Physicians should optimally meet face-to-face or by phone to ensure clear communication and thoughtful planning. It is important to note that patients often receive or perceive different messages from various care providers; thus, attempts should be made to provide consistency among these messages to the patient. Management plans and treatment options should be outlined succinctly and clearly for the patient.

Another dimension of team-based care involves allied health professions. It is not unusual for a hospitalized patient to encounter physical therapists, pharmacists, respiratory therapists, radiology technicians, social workers, dieticians, and transport personnel (among others) in addition to physicians and nurses. Each of these individuals contributes to clinical care as well as to the patient's experience with the health care system. In the outpatient setting, disease screening and chronic disease management are often carried out by nurses, physician assistants, or other allied health professionals.

The growth of team-based care has important implications for medical culture, student and resident training, and the organization of health care systems. Despite diversity in training, skills, and responsibilities among health care professionals, common values need to be espoused and reinforced. Many medical schools have incorporated interprofessional teamwork into their curricula. Effective communication is inevitably the most challenging aspect of implementing team-based care. While communication can be aided by electronic devices, including medical records, apps, or text messages, it is vitally important to balance efficiency with taking the necessary time to speak directly with colleagues.

The Dichotomy of Inpatient and Outpatient Internal Medicine The hospital environment has undergone sweeping changes over the past few decades. Emergency departments and critical care units have evolved to manage critically ill patients, allowing them to survive formerly fatal conditions. In parallel, there is increasing pressure to reduce the length of stay in the hospital and to manage complex disorders in the outpatient setting. This transition has been driven not only by efforts to reduce costs but also by the availability of new outpatient technologies, such as imaging and percutaneous infusion catheters for long-term antibiotics or nutrition, minimally invasive surgical procedures, and evidence that outcomes often are improved by reducing inpatient hospitalization.

In addition to traditional medical beds, hospitals now encompass multiple distinct levels of care, such as the emergency department, procedure rooms, overnight observation units, critical care units, and palliative care units. A consequence of this differentiation has been the emergence of new specialties (e.g., emergency medicine and end-of-life care) and the provision of in-hospital care by hospitalists and intensivists. Most *hospitalists* are board-certified internists who bear primary responsibility for the care of hospitalized patients and whose work is limited entirely to the hospital setting. The shortened length of hospital stay means that most patients receive only acute care while hospitalized; the increased complexities of inpatient medicine make the presence of an internist with specific training, skills, and experience in the hospital environment extremely beneficial. *Intensivists* are board-certified physicians who are further certified in critical care medicine and who direct and provide care for very ill patients in critical care units. Clearly, an important challenge in internal medicine today is to ensure the continuity of communication and information flow between a patient's primary care physician and those who are in charge of the patient's hospital care. Maintaining these channels of communication is frequently complicated by patient "handoffs"—i.e., transitions from the outpatient to the inpatient environment, from the critical care unit to a general medicine floor, from a medical to a surgical service and vice versa, from the hospital environment to the recently developed "home hospital" setting (for select patients with adequate home support), and from the hospital or home hospital to the outpatient environment.

The involvement of many care providers in conjunction with these transitions can threaten the traditional one-to-one relationship between patient and primary care physician. Of course, patients can benefit greatly from effective collaboration among a number of health care professionals; however, *it is the duty of the patient's principal or primary physician to provide cohesive guidance through an illness*. To meet this challenge, primary care physicians must be familiar with the techniques, skills, and objectives of specialist physicians and allied health professionals who care for their patients in the hospital. In addition, primary care physicians must ensure that their patients benefit from scientific advances and the expertise of specialists, both in and out of the hospital. Primary care physicians should explain the role of these specialists to reassure patients that they are in the hands of physicians best trained to manage their current illness. However, the primary care physician should assure patients and their families that decisions are being made in consultation with these specialists. The evolving concept of the "medical home" incorporates team-based primary care with subspecialty care in a cohesive environment that ensures smooth transitions of care.

Mitigating the Stress of Acute Illness Few people are prepared for a new diagnosis of cancer or anticipate the occurrence of a myocardial infarction, stroke, or major accident. The care of a frightened or distraught patient is confounded by these understandable responses to life-threatening events. The physician and other health providers can reduce the shock of life-changing events by providing information in a clear, calm, consistent, and reassuring manner. Often, information and reassurance need to be repeated. Caregivers should also recognize that, for the typical patient, hospital emergency rooms, operating rooms, ICUs, and general medical floors represent an intimidating environment. Hospitalized patients find themselves surrounded by air jets, buttons, and glaring lights; invaded by tubes and wires; and beset by the numerous members of the health care team—hospitalists, specialists, nurses, nurses' aides, physician assistants, social workers, technologists, physical therapists, medical students, house officers, attending and consulting physicians, and many others. They may be transported to special laboratories and imaging facilities replete with blinking lights, strange sounds, and unfamiliar personnel; they may be left unattended at times; and they may be obligated to share a room with other patients who have their own health problems. It is little wonder that patients may find this environment bewildering and stressful. The additive effects of an acute illness, unfamiliar environment, multiple medications, and sleep deprivation can lead to confusion or delirium, especially in older hospitalized patients. Physicians who appreciate the hospital experience from the patient's perspective and who make an effort to guide the patient through this experience may make a stressful situation more tolerable and enhance the patient's chances for an optimal recovery.

Medical Decision-Making Medical decision-making is a fundamental responsibility of the physician and occurs at each stage of the diagnostic and therapeutic process. The decision-making process involves the ordering of additional tests, requests for consultations, decisions about treatment, and predictions concerning prognosis. This process requires an in-depth understanding of the pathophysiology and natural history of disease. Formulating a differential diagnosis requires not only a broad knowledge base but also the ability to assess the relative probabilities of various diseases for a given patient. Application of the scientific method, including hypothesis formulation and data collection, is essential to the process of accepting or rejecting a particular diagnosis. Analysis of the differential diagnosis is an iterative process. As new information or test results are acquired, the group of disease processes being considered can be contracted or expanded appropriately. Whenever possible, decisions should be evidence-based, taking advantage of rigorously designed clinical trials or objective comparisons of different diagnostic tests. *Evidence-based medicine* stands in sharp contrast to anecdotal experience, which is often biased. Unless attuned to the importance of using larger, objective studies for making decisions, even the most experienced physicians can be influenced

to an undue extent by recent encounters with selected patients. Evidence-based medicine has become an increasingly important part of routine medical practice and has led to the publication of many useful practice guidelines. It is important to remember, however, that only a small fraction of the many decisions made in clinical practice are based on rigorous clinical trial evidence; other guideline recommendations are, therefore, predicated on expert consensus and weaker evidentiary support.

Thus, the importance of evidence-based medicine notwithstanding, much medical decision-making still relies on good clinical judgment, an attribute that is difficult to quantify or even to assess qualitatively. Physicians must use their knowledge and experience as a basis for weighing known factors, along with the inevitable uncertainties, and then making a sound judgment; this synthesis of information is particularly important when a relevant evidence base is not available. Several quantitative tools may be invaluable in synthesizing the available information, including diagnostic tests, Bayes' theorem (the probability of an event predicated on prior knowledge of conditions possibly related to the event), and multivariate statistical models ([Chap. 4](#)). Diagnostic tests serve to reduce uncertainty about an individual's diagnosis or prognosis and help the physician decide how best to manage that individual's condition. The battery of diagnostic tests complements the history and physical examination. The accuracy of a particular test is ascertained by determining its sensitivity (true-positive rate) and specificity (true-negative rate), as well as the predictive value of a positive and a negative result. [See Chap. 4 for a more thorough discussion of decision-making in clinical medicine.](#)

Practice Guidelines Many professional organizations and government agencies have developed formal clinical-practice guidelines to aid physicians and other caregivers in making diagnostic and therapeutic decisions that are evidence-based, cost-effective, and most appropriate to a particular patient and clinical situation. As the evidence base of medicine increases, guidelines can provide a useful framework for managing patients with particular diagnoses or symptoms. Clinical guidelines can protect patients—particularly those with inadequate health care benefits—from receiving substandard care. These guidelines also can protect conscientious caregivers from inappropriate charges of malpractice and society from the excessive costs associated with the overuse of medical resources. There are, however, caveats associated with clinical-practice guidelines since they tend to oversimplify the complexities of medicine. Furthermore, groups with different perspectives may develop divergent recommendations regarding issues as basic as the need for screening of women by mammography or of men with serum prostate-specific antigen (PSA) measurements. Finally, guidelines, as the term implies, do not—and cannot be expected to—account for the uniqueness of each individual and his or her illness. The physician's challenge is to integrate into clinical practice the useful recommendations offered by experts without accepting them blindly or being inappropriately constrained by them.

Precision Medicine The concept of *precision* or *personalized medicine* reflects the growing recognition that diseases once lumped together can be further stratified on the basis of genetic, biomarker, phenotypic, and/or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations. Inherent in this concept is the goal of targeting therapies in a more specific way to improve clinical outcomes for the individual patient and minimize unnecessary side effects for those less likely to respond to a particular treatment. In some respects, precision medicine represents the evolution of clinical practice guidelines, which are usually developed for populations of patients or a particular diagnosis (e.g., hypertension, thyroid nodule). As the pathobiology, prognosis, and treatment responses of subgroups within these diagnoses become better understood (i.e., through refined genomic analysis or enhanced deep phenotyping), the relevant clinical guidelines incorporate progressively more refined recommendations for individuals within these subgroups. The role of precision medicine is best illustrated for cancers in which genetic testing is able to predict responses (or the lack thereof) to

targeted therapies ([Chap. 73](#)). One can anticipate similar applications of precision medicine in pharmacogenomics, immunologic disorders, and diseases in which biomarkers can predict treatment responses. [See Chap. 5 for a more thorough discussion of precision medicine.](#)

Evaluation of Outcomes Clinicians generally use *objective* and readily measurable parameters to judge the outcome of a therapeutic intervention. These measures may oversimplify the complexity of a clinical condition as patients often present with a major clinical problem in the context of multiple complicating background illnesses. For example, a patient may present with chest pain and cardiac ischemia, but with a background of chronic obstructive pulmonary disease and renal insufficiency. For this reason, outcome measures, such as mortality, length of hospital stay, or readmission rates, are typically risk-adjusted. An important point to remember is that patients usually seek medical attention for *subjective* reasons; they wish to obtain relief from pain, to preserve or regain function, and to enjoy life. The components of a patient's health status or quality of life can include bodily comfort, capacity for physical activity, personal and professional function, sexual function, cognitive function, and overall perception of health. Each of these important domains can be assessed through structured interviews or specially designed questionnaires. Such assessments provide useful parameters by which a physician can judge patients' subjective views of their disabilities and responses to treatment, particularly in chronic illness. The practice of medicine requires consideration and integration of both objective and subjective outcomes.

Many health systems use survey and patient feedback data to assess qualitative features such as patient satisfaction, access to care, and communication with nurses and physicians. In the United States, HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) surveys are used by many systems and are publicly reported. Social media is also being used to assess feedback in real time as well as to share patient experiences with health care systems, potentially enriching the information available for use in medical decisions.

Errors in the Delivery of Health Care A series of reports from the Institute of Medicine (now the National Academy of Medicine [NAM]) called for an ambitious agenda to reduce medical error rates and improve patient safety by designing and implementing fundamental changes in health care systems ([Chap. 8](#)). It is the responsibility of hospitals and health care organizations to develop systems to reduce risk and ensure patient safety. Medication errors can be reduced through the use of ordering systems that rely on electronic processes or, when electronic options are not available, that eliminate misreading of handwriting. Whatever the clinical situation, it is the physician's responsibility to use powerful therapeutic measures wisely, with due regard for their beneficial actions, potential dangers, and cost. Implementation of infection control systems, enforcement of hand-washing protocols, and careful oversight of antibiotic use can minimize the complications of nosocomial infections. Central-line infection rates and catheter-associated urinary tract infections have been dramatically reduced at many centers by careful adherence of trained personnel to standardized protocols for introducing and maintaining central lines and urinary catheters, respectively. Rates of surgical infection and wrong-site surgery can likewise be reduced by the use of standardized protocols and checklists. Falls by patients can be minimized by judicious use of sedatives and appropriate assistance with bed-to-chair and bed-to-bathroom transitions. Taken together, these and other measures are saving thousands of lives each year.

Electronic Medical Records Both the growing reliance on computers and the strength of information technology now play central roles in medicine, including efforts to reduce medical errors. Laboratory data are accessed almost universally through computers. Many medical centers now have electronic medical records (EMRs), computerized order entry, and bar-coded tracking of medications. Some of these systems are interactive, sending reminders or warning of potential medical errors.

EMRs offer rapid access to information that is invaluable in enhancing health care quality and patient safety, including relevant data,

historical and clinical information, imaging studies, laboratory results, and medication records. These data can be used to monitor and reduce unnecessary variations in care and to provide real-time information about processes of care and clinical outcomes. Ideally, patient records are easily transferred across the health care system; however, technological limitations and concerns about privacy and cost continue to limit broad-based use of EMRs in many clinical settings.

For all of the advantages of EMRs, they can create distance between the physician and patient if care is not taken to preserve face-to-face contact. EMRs also require training and time for data entry. Many providers spend significant time entering information to generate structured data and to meet billing requirements. They may feel pressured to take short cuts, such as “cutting and pasting” parts of earlier notes into the daily record, thereby increasing the risk of errors. EMRs also structure information in a manner that disrupts the traditional narrative flow across time and among providers. These features, which may be frustrating for some providers, must be weighed against the advantages of ready access to past medical history, imaging, laboratory data, and consultant notes. Furthermore, the effort, time, and attention needed to maintain and utilize the EMR have led to a growing sense of dissatisfaction among physicians, lessening professional and personal well-being as a result. Clearly, this is an area of daily practice that requires improvement both for the delivery of safe and optimal care and physician wellness.

It is important to emphasize that information technology is merely a tool and can never replace the clinical decisions that are best made by the physician. Clinical knowledge and an understanding of a patient’s needs, supplemented by quantitative tools, still represent the best approach to decision-making in the practice of medicine.

THE PATIENT-PHYSICIAN RELATIONSHIP

The significance of the intimate personal relationship between physician and patient cannot be too strongly emphasized, for in an extraordinarily large number of cases both the diagnosis and treatment are directly dependent on it. One of the essential qualities of the clinician is interest in humanity, for the secret of the care of the patient is in caring for the patient.

—Francis W. Peabody, October 21, 1925,
Lecture at Harvard Medical School

Physicians must never forget that patients are individuals with problems that all too often transcend their physical complaints. They are not “cases” or “admissions” or “diseases.” Patients do not fail treatments; treatments fail to benefit patients. This point is particularly important in this era of high technology in clinical medicine. Most patients are anxious and fearful. Physicians should instill confidence and offer reassurance, but they must never come across as arrogant, patronizing, impatient, or hurried. A professional attitude, coupled with warmth and openness, can do much to alleviate anxiety and to encourage patients to share all aspects of their medical history. Empathy and compassion are the essential features of a caring physician. The physician needs to consider the setting in which an illness occurs—in terms not only of patients themselves but also of their familial, social, and cultural backgrounds. The ideal patient–physician relationship is based on thorough knowledge of the patient, mutual trust, and the ability to communicate.

Informed Consent The fundamental principles of medical ethics require physicians to act in the patient’s best interest and to respect the patient’s autonomy. Both principles are reflected in the process of informed consent. Patients are required to sign consent forms for most diagnostic or therapeutic procedures. Many patients possess limited medical knowledge and must rely on their physicians for advice. Communicating in a clear and understandable manner, physicians must fully discuss the alternatives for care and explain the risks, benefits, and likely consequences of each alternative. The physician is responsible for ensuring that the patient thoroughly understands these risks and benefits; encouraging questions is an important part of this process. It may be necessary to go over certain issues with the patient more than once. This is the very definition of *informed consent*. Complete, clear

explanation and discussion of the proposed procedures and treatment can greatly mitigate the fear of the unknown that commonly accompanies hospitalization. Often the patient’s understanding is enhanced by repeatedly discussing the issues in an unthreatening and supportive way, answering new questions that occur to the patient as they arise. Continuing efforts to educate the patient are essential. Patients are frequently inhibited from understanding by the fear of an uncertain future and potential impact of the illness on themselves and their families. Clear communication can also help alleviate misunderstandings in situations where complications of intervention occur. Special care should also be taken to ensure that a physician seeking a patient’s informed consent has no real or apparent conflict of interest.

Approach to Grave Prognoses and Death No circumstance is more distressing than the diagnosis of an incurable disease, particularly when premature death is inevitable. What should the patient and family be told? What measures should be taken to maintain life? What can be done to optimize quality of life?

Transparency of information, delivered in an appropriate manner, is essential in the face of a terminal illness. Even patients who seem unaware of their medical circumstances, or whose family members have protected them from diagnoses or prognoses, often have keen insights into their condition. They may also have misunderstandings that can lead to additional anxiety. The patient must be given an opportunity to speak with the physician and ask questions. A wise and insightful physician uses such open communication as the basis for assessing what the patient wants to know and when he or she wants to know it. On the basis of the patient’s responses, the physician can assess the most appropriate time and pace for sharing information. Ultimately, the patient must understand the expected course of the disease so that appropriate plans and preparations can be made. The patient should participate in decision-making with an understanding of the goal of treatment (palliation) and its likely effects. The patient’s religious beliefs should be taken into consideration. Some patients may find it easier to share their feelings about death with their physician, nurses, or members of the clergy than with family members or friends.

The physician should provide or arrange for emotional, physical, and spiritual support, and must be compassionate, unhurried, and open. In many instances, there is much to be gained by the laying on of hands. Pain should be controlled adequately, human dignity maintained, and isolation from family and close friends avoided. These aspects of care tend to be overlooked in hospitals, where the intrusion of life-sustaining equipment can detract from attention to the individual person and encourage concentration instead on the life-threatening disease, against which the battle ultimately will be lost in any case. In the face of terminal illness, the goal of medicine must shift from *cure* to *care* in the broadest sense of the term. *Primum succurrere*, first to help, is a guiding principle. In offering care to a dying patient, a physician should be prepared to provide information to family members and deal with their grief and sometimes their feelings of guilt or even anger. It is important for the physician to assure the family that everything reasonable is being done. A substantial challenge in these discussions is that the physician often does not know exactly how to gauge the prognosis. In addition, various members of the health care team may offer different opinions. Good communication among providers is essential so that consistent information is provided to patients. This is especially important when the best path forward is uncertain. Advice from experts in palliative and terminal care should be sought whenever appropriate to ensure that clinicians are not providing patients with unrealistic expectations. **For a more complete discussion of end-of-life care, see Chap. 12.**

Maintaining Humanism and Professionalism Many trends in the delivery of health care tend to make medical care impersonal. These trends, some of which have been mentioned already, include (1) vigorous efforts to reduce the escalating costs of health care; (2) the growing number of managed-care programs, which are intended to reduce costs but where the patient may have little choice in selecting a physician; (3) increasing reliance on technological advances and

computerization; and (4) the need for numerous physicians and other health professionals to be involved in the care of most patients who are seriously ill.

In light of these changes in the medical care system, it is a major challenge for physicians to maintain the *humane* aspects of medical care. The American Board of Internal Medicine, working together with the American College of Physicians–American Society of Internal Medicine and the European Federation of Internal Medicine, has published a *Charter on Medical Professionalism* that underscores three main principles in physicians' contract with society: (1) the primacy of patient welfare, (2) patient autonomy, and (3) social justice. While medical schools appropriately place substantial emphasis on professionalism, a physician's personal attributes, including integrity, respect, and compassion, also are extremely important. In the United States, the Gold Humanism Society recognizes individuals who are exemplars of humanistic patient care and serve as role models for medical education and training.

Availability to the patient, expression of sincere concern, willingness to take the time to explain all aspects of the illness, and a nonjudgmental attitude when dealing with patients whose cultures, lifestyles, attitudes, and values differ from those of the physician are just a few of the characteristics of a humane physician. Every physician will, at times, be challenged by patients who evoke strongly negative or positive emotional responses. Physicians should be alert to their own reactions to such situations and should consciously monitor and control their behavior so that the patient's best interest remains the principal motivation for their actions at all times.

Another important aspect of patient care involves an appreciation of the patient's "quality of life," a subjective assessment of what each patient values most. This assessment requires detailed, sometimes intimate knowledge of the patient, which usually can be obtained only through deliberate, unhurried, and often repeated conversations. Time pressures will always threaten these interactions, but they should not diminish the importance of understanding and seeking to fulfill the priorities of the patient.

■ EXPANDING FRONTIERS IN MEDICAL PRACTICE

The Era of "Omics" In the spring of 2003, announcement of the complete sequencing of the human genome officially ushered in the genomic era. However, even before that landmark accomplishment, the practice of medicine had been evolving as a result of insights into both the human genome and the genomes of a wide variety of microbes. The clinical implications of these insights are illustrated by the complete genome sequencing of H1N1 influenza virus in 2009 and even faster sequencing of COVID-19 in early 2020, leading to the swift development and dissemination of effective vaccines. Today, gene expression profiles are being used to guide therapy and inform prognosis for a number of diseases, and genotyping is providing a new means to assess the risk of certain diseases as well as variations in response to a number of drugs. Despite these advances, the use of complex genomics in the diagnosis, prevention, and treatment of disease is still in its early stages. The task of physicians is complicated by the fact that phenotypes generally are determined not by genes alone but by the complex interactions among genes and gene products, and by the interplay of genetic and environmental factors.

Rapid progress is also being made in other areas of molecular medicine. *Epigenetics* is the study of alterations in chromatin and histone proteins and methylation of DNA sequences that influence gene expression ([Chap. 483](#)). Every cell of the body has identical DNA sequences; the diverse phenotypes a person's cells manifest are, in part, the result of epigenetic regulation of gene expression. Epigenetic alterations are associated with a number of cancers and other diseases. *Proteomics*, the study of the entire library of proteins made in a cell or organ and the complex relationship of these proteins to disease, is enhancing the repertoire of the 23,000 genes in the human genome through alternate splicing, posttranslational processing, and posttranslational modifications that often have unique functional consequences. The presence or absence of particular proteins in the circulation or in cells is being explored for many diagnostic and disease-screening

applications. *Microbiomics* is the study of the resident microbes in humans and other mammals, which together compose the microbiome. The human haploid genome has ~23,000 genes, whereas the microbes residing on and in the human body encompass more than 3–4 million genes; these resident microbes are likely to be of great significance with regard to health status. Ongoing research is demonstrating that the microbes inhabiting human mucosal and skin surfaces play a critical role in maturation of the immune system, in metabolic balance, in brain function, and in disease susceptibility. A variety of environmental factors, including the use and overuse of antibiotics, have been tied experimentally to substantial increases in disorders such as obesity, metabolic syndrome, atherosclerosis, and immune-mediated diseases in both adults and children. *Metagenomics*, of which microbiomics is a part, is the genomic study of environmental species that have the potential to influence human biology directly or indirectly. An example is the study of exposures to microorganisms in farm environments that may be responsible for the lower incidence of asthma among children raised on farms. *Metabolomics* is the study of the range of metabolites in cells or organs and the ways they are altered in disease states. The aging process itself may leave telltale metabolic footprints that allow the prediction (and possibly the prevention) of organ dysfunction and disease. It seems likely that disease-associated patterns will be found in lipids, carbohydrates, membranes, mitochondria and mitochondrial function, and other vital components of cells and tissues. *Exposomics* is the study of the exposome—i.e., the environmental exposures such as smoking, sunlight, diet, exercise, education, and violence that together have an enormous impact on health. All of this new information represents a challenge to the traditional reductionist approach to medical thinking. The variability of results in different patients, together with the large number of variables that can be assessed, creates challenges in identifying preclinical disease and defining disease states unequivocally. Accordingly, the tools of *systems biology* and *network medicine* are being applied to the enormous body of information ("big data") now obtainable for every patient and may eventually provide new approaches to classifying disease. [For a more complete discussion of a complex systems and network science approach to human disease, see Chap. 486.](#)

The rapidity of these advances may seem overwhelming to practicing physicians; however, physicians have an important role to play in ensuring that these powerful technologies and sources of new information are applied judiciously to patient care. Since omics are evolving so rapidly, physicians and other health care professionals must engage in continuous learning so that they can apply this new knowledge to the benefit of their patients' health and well-being. Genetic testing requires wise counsel based on an understanding of the value and limitations of the tests as well as the implications of their results for specific individuals. [For a more complete discussion of genetic testing, see Chap. 467.](#)

The Globalization of Medicine Physicians should be cognizant of diseases and health care services beyond local boundaries. Global travel has critical implications for disease spread, and it is not uncommon for diseases endemic to certain regions to be seen in other regions after a patient has traveled to and returned from those regions. The outbreak of Zika virus infections in the Americas is a cogent example of this phenomenon. In addition, factors such as wars, the migration of refugees, and increasing climate extremes are contributing to changing disease profiles worldwide. Patients have broader access to unique expertise or clinical trials at distant medical centers, even those in other countries, and the cost of travel may be offset by the quality of care at those distant locations. As much as any other factor influencing global aspects of medicine, the Internet has transformed the transfer of medical information throughout the world. This change has been accompanied by the transfer of technological skills through telemedicine and international consultation—for example, interpretation of radiologic images and pathologic specimens. [For a complete discussion of global issues, see Chap. 472.](#)

Medicine on the Internet On the whole, the Internet has had a positive effect on the practice of medicine; through personal computers, a wide range of information is available to physicians and patients

almost instantaneously at any time and from anywhere in the world. This medium holds enormous potential for the delivery of current information, practice guidelines, state-of-the-art conferences, journal content, textbooks (including this text), and direct communications with other physicians and specialists, expanding the depth and breadth of information available to the physician regarding the diagnosis and care of patients. Medical journals are now accessible online, providing rapid sources of new information. By bringing them into direct and timely contact with the latest developments in medical care, this medium also serves to lessen the information gap that has hampered physicians and health care providers in remote areas.

Patients, too, are turning to the Internet in increasing numbers to acquire information about their illnesses and therapies and to join Internet-based support groups. Patients often arrive at a clinic visit with sophisticated information about their illnesses. In this regard, physicians are challenged in a positive way to keep abreast of the latest relevant information while serving as an “editor” as patients navigate this seemingly endless source of information, the accuracy and validity of which are not uniform.

A critically important caveat is that virtually anything can be published on the Internet, with easy circumvention of the peer-review process that is an essential feature of academic publications. Both physicians and patients who search the Internet for medical information must be aware of this danger. Notwithstanding this limitation, appropriate use of the Internet is revolutionizing information access for physicians and patients, and in this regard represents a remarkable resource that was not available to practitioners a generation ago.

Public Expectations and Accountability The general public’s level of knowledge and sophistication regarding health issues has grown rapidly over the past few decades. As a result, expectations of the health care system in general and of physicians in particular have risen. Physicians are expected to master rapidly advancing fields (the *science* of medicine) while considering their patients’ unique needs (the *art* of medicine). Thus, physicians are held accountable not only for the technical aspects of the care they provide but also for their patients’ satisfaction with the delivery and costs of care.

In many parts of the world, physicians increasingly are expected to account for the way in which they practice medicine by meeting certain standards prescribed by federal and local governments. The hospitalization of patients whose health care costs are reimbursed by the government and other third parties is subjected to utilization review. Thus, a physician must defend the cause for and duration of a patient’s hospitalization if it falls outside certain “average” standards. Authorization for reimbursement increasingly is based on documentation of the nature and complexity of an illness, as reflected by recorded elements of the history and physical examination. A growing “pay-for-performance” movement seeks to link reimbursement to quality of care. The goal of this movement is to improve standards of health care and contain spiraling health care costs. In many parts of the United States, managed (capitated) care contracts with insurers have replaced traditional fee-for-service care, placing the onus of managing the cost of all care directly on the providers and increasing the emphasis on preventive strategies. In addition, physicians are expected to give evidence of their current competence through mandatory continuing education, patient record audits, maintenance of certification, and relicensing.

Medical Ethics and New Technologies The rapid pace of technological advances has profound implications for medical applications that go far beyond the traditional goals of disease prevention, treatment, and cure. Cloning, genetic engineering, gene therapy, human-computer interfaces, nanotechnology, and use of targeted therapies have the potential to modify inherited predispositions to disease, select desired characteristics in embryos, augment “normal” human performance, replace failing tissues, and substantially prolong life span. Given their unique training, physicians have a responsibility to help shape the debate on the appropriate uses of and limits placed on these new technologies and to consider carefully the ethical issues associated with the implementation of such interventions. As medicine becomes more complex, shared decision-making is increasingly important, not

only in areas such as genetic counseling and end-of-life care, but also in diagnostic and treatment options.

Learning Medicine More than a century has passed since the publication of the Flexner Report, a seminal study that transformed medical education and emphasized the scientific foundations of medicine as well as the acquisition of clinical skills. In an era of burgeoning information and access to medical simulation and informatics, many schools are implementing new curricula that emphasize lifelong learning and the acquisition of competencies in teamwork, communication skills, system-based practice, and professionalism. The tools of medicine also change continuously, necessitating formal training in the use of EMRs, large datasets, ultrasound, robotics, and new imaging techniques. These and other features of the medical school curriculum provide the foundation for many of the themes highlighted in this chapter and are expected to allow physicians to progress, with experience and learning over time, from competency to proficiency to mastery.

At a time when the amount of information that must be mastered to practice medicine continues to expand, increasing pressures both within and outside of medicine have led to the implementation of restrictions on the amount of time a physician-in-training can spend in the hospital and in clinics. Because the benefits associated with continuity of medical care and observation of a patient’s progress over time were thought to be outstripped by the stresses imposed on trainees by long hours and by fatigue-related errors, strict limits were set on the number of patients that trainees could be responsible for at one time, the number of new patients they could evaluate in a day on call, and the number of hours they could spend in the hospital. In 1980, residents in medicine worked in the hospital more than 90 hours per week on average. In 1989, their hours were restricted to no more than 80 per week. Resident physicians’ hours further decreased by ~10% between 1996 and 2008, and in 2010, the Accreditation Council for Graduate Medical Education further restricted (i.e., to 16 hours per shift) consecutive in-hospital duty hours for first-year residents. The impact of these changes is still being assessed, but the evidence that medical errors have decreased as a consequence is sparse. An unavoidable by-product of fewer hours at the bedside is an increase in the number of “handoffs” of patient responsibility from one physician to another. These transfers often involve a transition from a physician who knows the patient well, having evaluated that individual on admission, to a physician who knows the patient less well. It is imperative that these transitions of responsibility be handled with care and thoroughness, with all relevant information exchanged and acknowledged. These issues highlight the challenge our profession has in establishing a reliable measure of physician effectiveness.

The Physician as Perpetual Student From the time physicians graduate from medical school, it becomes all too apparent that this milestone is symbolic and that they must embrace the role of a “perpetual student.” This realization is at the same time exhilarating and anxiety-provoking. It is exhilarating because physicians can apply constantly expanding knowledge to the treatment of their patients; it is anxiety-provoking because physicians realize that they will never know as much as they want or need to know. Ideally, physicians will translate the latter feeling into energy through which they can continue to improve and reach their potential. It is the physician’s responsibility to pursue new knowledge continually by reading, attending conferences and courses, and consulting colleagues and the Internet. This is often a difficult task for a busy practitioner; however, a commitment to continued learning is an integral part of being a physician and must be given the highest priority.

The Physician as Citizen Being a physician is a privilege. The capacity to apply one’s skills for the benefit of fellow human beings is a noble calling. The physician-patient relationship is inherently unbalanced in the distribution of power. In light of their influence, physicians must always be aware of the potential impact of what they do and say, and must always strive to strip away individual biases and preferences to find what is best for their patients. To the extent possible, physicians should also act within their communities to promote health

and alleviate suffering. Meeting these goals begins by setting a healthy example and continues in taking action to deliver needed care even when personal financial compensation may not be available.

Research, Teaching, and the Practice of Medicine The word *doctor* is derived from the Latin *docere*, “to teach.” As teachers, physicians should share information and medical knowledge with colleagues, students of medicine and related professions, and their patients. The practice of medicine is dependent on the sum total of medical knowledge, which in turn is based on an unending chain of scientific discovery, clinical observation, analysis, and interpretation. Advances in medicine depend on the acquisition of new information through research, and improved medical care requires the transmission of that information. As part of their broader societal responsibilities, physicians should encourage patients to participate in ethical and properly approved clinical investigations if these studies do not impose undue hazard, discomfort, or inconvenience. Physicians engaged in clinical research must be alert to potential conflicts of interest between their research goals and their obligations to individual patients. The best interests of the patient must always take priority.

To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease—these are our ambitions.

—William Osler, 1849–1919

FURTHER READING

- CHESTON CC et al: Social media use in medical education: A systematic review. *Acad Med* 88:893, 2013.
- COOKE M et al: American medical education 100 years after the Flexner report. *N Engl J Med* 355:1339, 2006.
- EXCEL JL et al: Vaccine development for emerging infectious diseases. *Nat Med* 27:591, 2021.
- INSTITUTE OF MEDICINE: *Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life*. Washington, DC, National Academies Press, 2015.
- INSTITUTE OF MEDICINE: *Improving Diagnosis in Health Care*. Washington, DC, National Academies of Sciences, Engineering, and Medicine, 2015.
- LEVINE DM et al: Hospital-level care at home for acutely ill adults: A qualitative evaluation of a randomized controlled trial. *J Gen Intern Med* 36:1965, 2021.
- STERN DT, PAPADAKIS M: The developing physician—becoming a professional. *N Engl J Med* 355:1794, 2006.
- VICKREY BG et al: How neurologists think: A cognitive psychology perspective on missed diagnoses. *Ann Neurol* 67:425, 2010.
- WEST P et al: Intervention to promote physician well-being, job satisfaction, and professionalism. A randomized clinical trial. *JAMA Intern Med* 174:527, 2014.

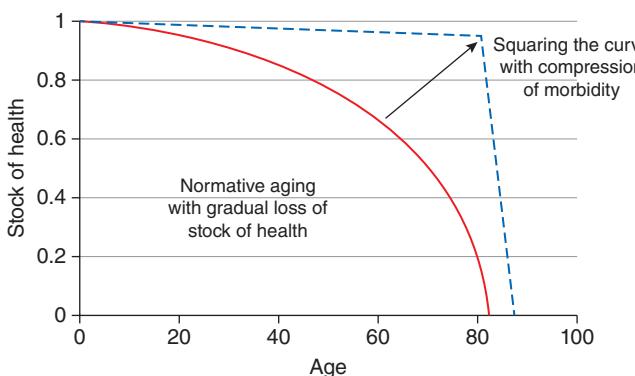


FIGURE 2-1 Loss of health with aging. Representation of normative aging with loss of the full stock of health with which individuals are born (indicating gain of morbidity), contrasted with a squared curve with greater longevity and fuller stock of health (less morbidity) until shortly before death. The “squared curve” represents the likely ideal situation for most patients.

worldwide over the last century (largely as a result of public health practices), increasing emphasis is placed on prevention for the purpose of preserving quality of life and extending the health span, not just the life span. Given that all patients will eventually die, the goal of prevention ultimately becomes compression of morbidity toward the end of the life span; that is, reduction of the amount of burden and time spent with disease prior to dying. As shown in Fig. 2-1, normative aging tends to involve a steady decline in the stock of health, with accelerating decline over time. Successful prevention offers the opportunity both to extend life and to extend healthy life, thus “squaring the curve” of health loss during aging.

Prevention strategies have been characterized as tertiary, secondary, primary, and primordial. *Tertiary prevention* requires rapid action to prevent imminent death in the setting of acute illness, such as through percutaneous coronary intervention in the setting of ST-segment elevation myocardial infarction. *Secondary prevention* strategies focus on avoiding the recurrence of disease and death in an individual who is already affected. For example, tamoxifen is recommended for women with surgically treated early-stage, estrogen receptor-positive breast cancer, because it reduces the risk of recurrent breast cancer (including in the contralateral breast) and death. *Primary prevention* attempts to reduce the risk of incident disease among individuals with one or more risk factors. Treatment of elevated blood pressure in individuals who have not yet experienced cardiovascular disease represents one example of primary prevention that has proven effective in reducing the incidence of stroke, heart failure, and coronary heart disease.

Primordial prevention is a more recent concept (first introduced in 1979) that focuses on prevention of the development of *risk factors* for disease, not just prevention of disease. Primordial prevention strategies emphasize upstream determinants of risk for chronic diseases, such as eating patterns, physical activity, and environmental and social determinants of health. It therefore encompasses medical treatment strategies for some individuals as well as a strong reliance on public health and social policy. It is increasingly clear that primordial prevention represents the ultimate means for reducing the burden of chronic diseases of aging. Once risk factors develop, it is difficult to restore risk to the low level of someone who never developed the risk factor. The time spent with adverse levels of the risk factor often causes irreversible damage that precludes complete restoration of low risk. For example, individuals with hypertension who are treated back to optimal levels ($<120/<80$ mmHg) do have a lower risk compared with untreated patients with hypertension, but they still have twice the risk of cardiovascular events as those who maintained optimal blood pressure without medications. Patients with elevated blood pressure that is subsequently treated have greater left ventricular mass index, worse renal function, and more evidence of atherosclerosis and other target organ damage as a result of the time spent with elevated blood pressure; such damage cannot be fully reversed despite efficacious therapy with antihypertensive medications. Conversely, as described below in greater detail, individuals who maintain optimal levels of all major

2

Promoting Good Health

Donald M. Lloyd-Jones,
Kathleen M. McKibbin

GOALS AND APPROACHES TO PREVENTION

Prevention of acute and chronic diseases before their onset has been recognized as one of the hallmarks of excellent medical practice for centuries and is now used as a metric for highly functioning health care systems. The ultimate goal of preventive strategies is to avoid premature death. However, as longevity has increased dramatically

cardiovascular risk factors into middle age through primordial prevention essentially abolish their lifetime risk of developing cardiovascular disease while also living substantially longer and having a lower burden and later onset of other comorbid illnesses (compression of morbidity).

Prevention strategies should be distinguished from disease screening strategies. Screening attempts to detect evidence of disease at its earliest stages, when treatment is likely to be more efficacious than for advanced disease (*Chap. 6*). Screening can be performed in service of prevention, especially if it aids in identifying preclinical markers, such as dyslipidemia or hyperglycemia, associated with elevated disease risk.

■ HEALTH PROMOTION

In recent decades, medical practice has increasingly focused on clinical and public health approaches to promote health, and not just prevent disease. Prevention of disease is a worthy individual and societal goal in and of itself, but it does not necessarily guarantee health. Health is a broader construct encompassing more than just absence of disease. It includes biologic, physiologic, and psychological domains (among others) in a continuum, rather than occurring as a dichotomous trait. Health is therefore somewhat subjective, but attempts have been made to use more objective criteria to define health in order to raise awareness, prevent disease, and promote healthy longevity.

For example, in 2010, the American Heart Association (AHA) defined a new construct of “cardiovascular health” based on evidence of associations with longevity, disease avoidance, healthy longevity, and quality of life. The definition of cardiovascular health is based on seven health behaviors and health factors (eating pattern, physical activity, smoking status, body mass index [BMI], and levels of blood pressure, blood cholesterol, and blood glucose) and includes a spectrum from poor to ideal. Individuals with optimal levels of all seven metrics simultaneously are considered to have ideal cardiovascular health. The state of cardiovascular health for an individual or a population can be assessed with simple scoring by counting the number of ideal metrics (out of seven) or applying 0 points for each poor metric, 1 point for each intermediate metric, and 2 points for each ideal metric, thus creating a composite cardiovascular health score ranging from 0 to 14 points. Higher cardiovascular health scores in younger and middle ages have been associated with greater longevity, lower incidence of cardiovascular disease, lower incidence of other chronic diseases of aging (including dementia, cancer, and more), compression of morbidity, greater quality of life, and lower health care costs, achieving both individual and societal goals for healthy aging and further establishing the critical importance of primordial prevention and cardiovascular health promotion.

Focusing on health promotion, rather than just disease prevention, may also provide greater motivation for patients to pursue lifestyle changes or adhere to clinician recommendations. Extensive literature suggests that providing patients solely with information regarding disease risk, or risk reduction with treatment, is unlikely to motivate desired behavior change. Empowering patients with strategies to achieve positive health goals after discussing risks can provide more effective adherence and better long-term outcomes. In the case of smoking cessation, enumerating only the risks of smoking can lead to patient inertia and therapeutic nihilism and has proven to be an ineffective approach, whereas strategies that incorporate positive health messaging, support, and feedback, with appropriate use of evidence-based therapies, have proven far more effective.

■ PRIORITIZING PREVENTION STRATEGIES

In secondary prevention, the patient already has manifest clinical disease and is therefore at high risk for progression. The approach should be to work with the patient to implement all evidence-based strategies that will help to prevent recurrence or progression. This will typically include drug therapy as well as therapeutic lifestyle changes to control ongoing risk factors that may have caused disease in the first place. Juggling priorities can be difficult, and barriers to implementation are many, including costs, time, patient health literacy, and patient and caregiver capacity to organize the regimen. Addressing these potential barriers with the patient can help to forge a therapeutic bond and

may improve adherence; ignoring them will likely lead to therapeutic failure. Numerous studies demonstrate that, even in high-functioning health systems, only ~50% of patients are taking recommended, evidence-based secondary prevention medications, such as statins, by 1 year after a myocardial infarction.

In patients who are eligible for primary prevention strategies, it is important to frame the discussion around the overall evidence base as well as an individual patient’s likelihood of benefit from a given preventive intervention. A first step is to understand the patient’s estimated absolute risk for disease in the foreseeable future or during their remaining life span. However, absolute risk estimation and presentation of those risks are generally insufficient to motivate behavior change. It is critical to assess the patient’s understanding and tolerance of the risk, their readiness to implement lifestyle changes or adhere to drug therapy, and their overall preferences regarding use of drug therapy to prevent an event (e.g., cancer, myocardial infarction, stroke). The clinician can help the patient by informing them of the risks for disease and potential for absolute benefits (and harms) from the available evidence-based choices. This may take more than one conversation, but given that diseases, such as cancer and cardiovascular disease, are the leading causes of premature death and disability, the time is well spent.

Partnering with the patient through motivational interviewing may assist in the process of selecting initial approaches to prevention. Selecting an area that the patient feels they are ready to change can lead to better adherence and greater achievement of success in the short and longer term. If the patient is uncertain what course to choose, prudence would dictate focusing on control of risk factors that may lead to the most rapid reduction in risk for acute events. For example, blood pressure is both a chronic risk factor and an acute trigger for cardiovascular events. Thus, if a patient has both significant elevations in blood pressure and dyslipidemia, it would be appropriate to focus initial efforts on blood pressure control. Likewise, a focus on smoking cessation can lead to more rapid reductions in risk for acute events than some other lifestyle interventions.

■ PREVENTION AND HEALTH PROMOTION ACROSS THE LIFE COURSE

Periodic Health Evaluations The “routine annual physical” has in many ways become an expected part of the patient-physician relationship in primary care practice. However, evidence for the efficacy of the periodic health evaluation in asymptomatic adults unselected for risk factors or disease is mixed and depends on the outcome. Systematic reviews and meta-analyses of published trials have consistently observed lack of benefit (and also lack of harm) in terms of total mortality in association with periodic health evaluations. Data are more heterogeneous but overall suggest no benefit for cancer- or cardiovascular-specific mortality, with the potential for either benefit or harm depending on number of evaluations and patient-level factors. Well-designed studies on nonfatal clinical events and morbidity have been sparsely reported, but there appear to be no large effects.

Periodic health evaluations do appear to lead to greater diagnosis of certain conditions such as hypertension and dyslipidemia, as expected. Likewise, periodic health examinations also improve the delivery of recommended preventive services, such as gynecologic examinations and Papanicolaou smears, fecal occult blood testing, and cholesterol screening. The benefits and risks associated with screening tests are discussed in detail in *Chap. 6*. Risks of routine evaluations include inappropriate testing or overtesting or false-positive findings that require follow-up and induce patients to worry. Periodic health examinations appear to be associated with less patient worry. On balance, given the lack of convincing evidence of harm and the potential for better delivery of appropriate screening, counseling, and preventive services, periodic health evaluations appear reasonable for general populations at average risk for chronic conditions.

It is important to note that routine annual comprehensive physical examinations of asymptomatic adult patients have very low yield and may take an inordinate amount of time in a wellness visit. Such time

TABLE 2-1 Guidelines and Key Recommendations from the *Dietary Guidelines for Americans, 2020–2025*

GUIDELINES	KEY RECOMMENDATIONS
<p>1. Follow a healthy dietary pattern at every life stage. For the first 6 months of life, infants should exclusively be fed human milk, or iron-fortified formula if human milk is unavailable. From 6 to 12 months, infants should be introduced to a variety of complementary nutrient-dense foods. From 12 months to older adulthood, the dietary pattern should meet nutrient needs, help achieve a healthy body weight, and reduce the risk of chronic disease.</p> <p>2. Customize and enjoy nutrient-dense food and beverage choices to reflect personal preferences, cultural traditions, and budgetary considerations. The Dietary Guidelines provide a framework of several dietary patterns intended to be customized to individual needs and preferences, as well as the foodways of the diverse cultures in the United States.</p> <p>3. Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits. Nutrient-dense foods provide vitamins, minerals, and other health-promoting components and have no or little added sugars, saturated fat, and sodium. A healthy dietary pattern consists of nutrient-dense forms of foods and beverages across all food groups, in recommended amounts, and within calorie limits.</p> <p>4. Limit foods and beverages higher in added sugars, saturated fat, and sodium, and limit alcoholic beverages. At every life stage, meeting food group recommendations, even with nutrient-dense choices, fulfills most of a person's daily calorie needs and sodium limits, with little room for extra added sugars, saturated fat, or sodium, or for alcoholic beverages.</p>	<p>The Dietary Guidelines' Key Recommendations for healthy eating patterns should be applied in their entirety, given the interconnected relationship that each dietary component can have with others. They are also intended as a framework to accommodate personal preferences, cultural traditions, and budgetary considerations.</p> <p>Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits to achieve a healthy weight and reduce the risk of chronic disease.</p> <p>The core elements that make up a healthy dietary pattern include:</p> <ul style="list-style-type: none"> • Vegetables of all types—dark green; red and orange; beans, peas, and lentils; starchy; and other vegetables • Fruits, especially whole fruit • Grains, at least half of which are whole grain • Dairy, including fat-free or low-fat milk, yogurt, and cheese, and/or lactose-free versions and fortified soy beverages and yogurt as alternatives • Protein foods, including lean meats, poultry, and eggs; seafood; beans, peas, and lentils; and nuts, seeds, and soy products • Oils, including vegetable oils and oils in food, such as seafood and nuts <p>A healthy eating pattern limits:</p> <ul style="list-style-type: none"> • Added sugars—Less than 10% of calories per day starting at age 2. Avoid foods and beverages with added sugars for those younger than age 2. • Saturated fat—Less than 10% of calories per day starting at age 2. • Sodium—Less than 2300 mg per day—and even less for children younger than age 14. • Alcoholic beverages—Adults of legal drinking age can choose not to drink or to drink in moderation by limiting intake to 2 drinks or less in a day for men and 1 drink or less in a day for women, when alcohol is consumed. Drinking less is better for health than drinking more. There are some adults who should not drink alcohol, such as women who are pregnant. <p>Meet the U.S. Department of Health and Human Services' <i>Physical Activity Guidelines for Americans</i></p> <p>In tandem with the recommendations above, Americans of all ages—children, adolescents, adults, and older adults—should meet the <i>Physical Activity Guidelines for Americans</i> to help promote health and reduce the risk of chronic disease. Americans should aim to achieve and maintain a healthy body weight. The relationship between diet and physical activity contributes to calorie balance and managing body weight.</p>

Source: Adapted from the *Dietary Guidelines for Americans, 2020–2025*. Washington, DC: U.S. Department of Agriculture and U.S. Department of Health and Human Services; 2020. Available at https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf.

may be better spent on assessing and counseling the patient on other aspects of their health, as discussed below. Evidence-based components that should be included in periodic evaluations focused on health and prevention include a number of age-appropriate screening tests for chronic disease and risk factors, preventive interventions including immunizations and chemoprevention for at-risk individuals, and preventive counseling. The U.S. Preventive Services Task Force publishes its *Guide to Clinical Preventive Services*, which contains evidence-based recommendations from the Task Force on preventive services for which there is a high degree of certainty that the service provides at least moderate net clinical benefit (i.e., benefits outweigh harms significantly and to a reasonable magnitude).

Healthy Behaviors and Lifestyles Owing to the paucity of evidence, the heterogeneity of study designs, and the diverse nature of interventions studied, many clinicians are uncertain as to how to deliver advice regarding healthy behaviors and lifestyles. Nevertheless, adverse behaviors and lifestyles contribute to >75% of premature, preventable deaths and disability. Estimates from the U.S. National Health and Nutrition Examination Survey indicate that fewer than 1% of Americans achieve an optimal heart-healthy eating pattern. Thus, whereas there are many demands on time during a typical patient-clinician encounter, few things may have more impact on longevity, health, and quality of life for asymptomatic patients than an efficient approach to assessing, documenting, and improving patients' health behaviors. Indeed, the mere act of assessing health behaviors has been shown to affect patients' health behaviors. Facility with tools for assessment of lifestyle and with strategies for counseling are therefore of paramount importance.

Healthy Eating Patterns (see Chap. 332) Despite the existence of numerous "fad" diets and seemingly inconsistent recommendations

on dietary composition, there is remarkable agreement about what should constitute a healthy eating pattern for the broad population to avoid nutritional deficits (i.e., vitamin deficiency) and excesses (i.e., excessive caloric intake) and to maximize potential health (**Table 2-1**). Optimal eating patterns consist of whole fruits and vegetables, whole grains, lean proteins, and healthy oils, and allow for nonfat or low-fat dairy intake. They tend to exclude frequent ingestion of foods high in refined sugars and starches, saturated fat, and sodium. Since sodium and refined sugars and starches are the hallmark of much of the processed/packaged food supply, a simple rule of thumb is to provide or cook the majority of one's own meals starting from whole foods and emphasizing fruits and vegetables. Likewise, foods prepared outside of the home tend to have higher fat and sodium content, so special attention to menu choices focused on fruits, vegetables, lean proteins, and whole grains, while minimizing sauces and dressings, can help most individuals follow healthier eating patterns when eating food prepared outside the home. In all cases, sugar-sweetened beverages and nonnutritious snack foods should be minimized. If snacks are included, small amounts of healthy nuts and seeds or more fruits and vegetables should be encouraged.

Specific conditions and diseases, such as diabetes, other metabolic disorders, allergies, and gastrointestinal disorders, may require tailored approaches to diet. In counseling most patients, the general approach should focus on whole foods, eating patterns, and appropriate calorie balance, rather than on specific micronutrients such as electrolytes or selected vitamins. It should be remembered that most patients have difficulty understanding nutritional labels on packaged foods, with the attendant demands on numeracy and health literacy.

Dietary guidelines are published by the U.S. Department of Agriculture (USDA) and U.S. Department of Health and Human Services every 5 years, and these guidelines have undergone substantial

evolution over time. The current U.S. Dietary Guidelines and Key Recommendations for 2020–2025 are summarized in Table 2-1 and emphasize the importance of healthy eating patterns for every stage of life, to avoid chronic diseases including obesity, diabetes, cancer, and cardiovascular disease. The core elements include eating patterns with nutrient-dense (rather than calorie-dense) whole foods and appropriate caloric intake to achieve and maintain healthy weight. The USDA guidelines focus on the concept of a healthy plate (rather than the prior food pyramid) for ease of counseling and adoption. Fifty percent of the plate should consist of vegetables and whole fruits, with remaining portions for whole grains and lean protein foods. When using fat for cooking, it should be done by sauteing in healthier oils (e.g., canola oil), and addition of judicious amounts of healthy raw oils (e.g., olive oil, nuts) to dishes is appropriate. Recommendations also focus on limitation of foods and beverages higher in added sugars, saturated fat, and sodium, and moderation or avoidance of alcohol intake.

The USDA guidelines focus on specific healthy eating patterns that adhere to these broad recommendations and are appropriate for ~97% of the general population. They identify a “Healthy U.S.-Style Dietary Pattern” that adheres closely to the evidence-based Dietary Approaches to Stop Hypertension (DASH) eating pattern but is customizable for different cultural or personal preferences. Alternative patterns, which vary more in emphasis than in content, include a “Healthy Mediterranean-Style Dietary Pattern” and a “Healthy Vegetarian Dietary Pattern.”

AGE- AND SEX-SPECIFIC RECOMMENDATIONS Current dietary framework recommendations are generally similar for all life stages from ages ≥ 12 months, but recommended levels of caloric intake (and hence amounts of foods) differ by age, sex, and physical activity level. For example, recommended caloric intake ranges from 1000 calories/d for sedentary 2-year-old children to as high as 3200 calories/d for active 16- to 18-year-old young men. Recommended caloric intakes peak in late adolescence or early adulthood for men and women and gradually decrease over ensuing decades.

As with all lifestyle counseling aimed at behavior change, dietary approaches that partner with the patient and utilize motivational interviewing strategies and shared goals and commitments tend to work best, as described below (see “Approach to the Patient”).

Physical Activity Similar to the approach to counseling regarding healthy eating patterns, recommendations on participation in physical activity emphasize the point that any physical activity is better than none. A simple rule of thumb for patients is: “If you are doing nothing, do something; and if you are doing something, do more, every day.” The evidence base for physical activity indicates that the marginal benefits from physical activity are greatest in advancing from no activity to low levels of moderate activity. With increasing duration and intensity of activity, there is a continued curvilinear increase in health benefits, but the marginal gains for each additional minute of moderate-to-vigorous activity slowly diminish. Thus, for adults, the recommended amount of physical activity is 150 min of moderate-intensity or 75 min of vigorous-intensity aerobic activity per week, performed in episodes of at least 5 min, and preferably spread throughout the week, plus participation in muscle-strengthening activity at least 2 days per week. Additional health benefits can be realized by engaging in physical activity beyond this amount.

In counseling patients regarding physical activity, it is important to note that sedentary time (e.g., seated at work or at home in front of electronic screens) has adverse health consequences independent of the lack of physical activity during these episodes. Therefore, even modest efforts like standing at the desk and doing gentle stretching for periods during the day may be beneficial. It is also important to emphasize that participating in a variety of aerobic activities (biking, swimming, walking, jogging, rowing, elliptical training, stair-climbing, etc.) can be beneficial and may help to avoid overuse injuries and boredom with the exercise regimen. If patients choose to participate in muscle-strengthening activities for health improvement, emphasis should be placed on weights that allow more repetitions (e.g., 3 sets of 15–20 repetitions that can be performed comfortably, with a rest period in between) and on avoiding breath-holding and straining against a closed glottis.

SUDDEN CARDIAC DEATH RISK Patients may express concerns regarding the risk of sudden cardiac death during exercise. Whereas the risk of sudden death during exercise does increase directly with the amount of time spent exercising, this association is substantially mitigated by training effects. Thus, patients embarking on an exercise program should be encouraged to increase the duration of aerobic exercise gradually as tolerated, aiming for episodes of at least 30 min 5 times a week as an ideal. Once a comfortable duration is reached, incorporating interval training periods of more intensive activity interspersed during the exercise can provide greater fitness gains.

EXTREME ENDURANCE ACTIVITIES As with other forms of exercise, extreme endurance activities such as triathlons and marathons should be undertaken only with appropriate and graded training. Such activities tend to take a greater toll on the musculoskeletal system over time than less extreme activities, and they are also associated with measurable damage to the myocardium and greater risks for other organ damage. Athletes participating in endurance activities routinely have elevations in cardiac troponin (a specific circulating marker of myocardial cell damage and death) at the end of the race, although elevations are lower in those who are well trained. Patients and clinicians should consider the patient’s overall health, specific limitations, potential for injury, and ability to train in decision-making regarding participation in endurance events.

AGE-SPECIFIC RECOMMENDATIONS The U.S. Department of Health and Human Services’ *Physical Activity Guidelines for Americans*, second edition (2018) (**Table 2-2**), recommend that preschool-aged children (aged 3–5 years) should be physically active throughout the day in a variety of activity types to enhance growth and development. Children and adolescents aged 6–17 years should participate in ≥ 60 min of physical activity daily, most of which should be moderate- or vigorous-intensity aerobic activity, including vigorous, muscle-strengthening, and bone-strengthening activities at least 3 days a week each. As noted above, adults aged 18–64 years are recommended to pursue at least 150 min of moderate-intensity or 75 min of vigorous-intensity aerobic activity per week (or equivalent combinations), with at least 2 days of muscle-strengthening activities. Adults aged ≥ 65 years should follow the adult guidelines or be as active as possible as abilities and conditions allow. For older adults, special emphasis is also placed on multicomponent physical activity that includes balance training as well as aerobic and muscle-strengthening activities.

Sleep Hygiene Sleeping between 7 and 9 h per night appears to be optimal for health in adults aged ≥ 18 years. Sleeping < 7 h is associated with adverse outcomes, including obesity, diabetes, elevated blood pressure, cardiovascular disease, depression, and all-cause mortality, as well as physiologic disturbances such as impaired immune function, increased pain sensitivity, and impaired cognitive performance. Conversely, achieving appropriate levels of sleep is associated with more success in weight loss, better blood pressure control among patients with hypertension, and improved mental health and performance. Regular sleep more than 9 h per night is appropriate for children and adolescents or individuals recovering from sleep deprivation or illness, but for most individuals, the effects on health are uncertain.

Patients often express concerns about the quantity and quality of their sleep. With aging, both aspects of sleep tend to decline, even without overt sleep disorders. Documentation of sleep using a sleep log may assist in understanding different types of insomnia and sleep disorders. Encouraging daily activity to promote fatigue, avoidance of eating and drinking alcohol too close to bedtime, and regular daily sleep habits may help patients achieve better sleep. Regular use of sedative medications should generally be discouraged given the high potential for dependence, addiction, and altered sleep quality.

DISORDERS OF SLEEP The prevalence of sleep-related breathing disorders, including obstructive sleep apnea (OSA), is poorly documented. A recent systematic review suggested that the prevalence of clinically important OSA in the general adult population may be between 9% and 38%, with higher rates in men versus women, older versus younger adults, and those with higher versus lower BMI.

TABLE 2-2 Recommendations from Physical Activity Guidelines for Americans

AGE	RECOMMENDATIONS
3–5 years	<ul style="list-style-type: none"> Preschool-aged children (ages 3 through 5 years) should be physically active throughout the day to enhance growth and development. Adult caregivers of preschool-aged children should encourage active play that includes a variety of activity types.
6–17 years	<ul style="list-style-type: none"> It is important to provide young people opportunities and encouragement to participate in physical activities that are appropriate for their age, that are enjoyable, and that offer variety. Children and adolescents ages 6 through 17 years should do 60 min (1 h) or more of moderate-to-vigorous physical activity daily: <ul style="list-style-type: none"> Aerobic: Most of the 60 min or more per day should be either moderate- or vigorous-intensity aerobic physical activity and should include vigorous-intensity physical activity on at least 3 days a week. Muscle-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include muscle-strengthening physical activity on at least 3 days a week. Bone-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include bone-strengthening physical activity on at least 3 days a week.
18–64 years	<ul style="list-style-type: none"> Adults should move more and sit less throughout the day. Some physical activity is better than none. Adults who sit less and do any amount of moderate-to-vigorous physical activity gain some health benefits. For substantial health benefits, adults should do at least 150 min (2 h and 30 min) to 300 min (5 hours) a week of moderate-intensity or 75 min (1 h and 15 min) to 150 min (2 h and 30 min) a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic activity. Preferably, aerobic activity should be spread throughout the week. Additional health benefits are gained by engaging in physical activity beyond the equivalent of 300 min (5 h) of moderate-intensity physical activity a week. Adults should also do muscle-strengthening activities of moderate or greater intensity and that involve all major muscle groups on 2 or more days a week, as these activities provide additional health benefits.
≥65 years	<ul style="list-style-type: none"> The key guidelines for adults also apply to older adults. In addition, the following key guidelines are just for older adults: <ul style="list-style-type: none"> As part of their weekly physical activity, older adults should do multicomponent physical activity that includes balance training as well as aerobic and muscle-strengthening activities. Older adults should determine their level of effort for physical activity relative to their level of fitness. Older adults with chronic conditions should understand whether and how their conditions affect their ability to do regular physical activity safely. When older adults cannot do 150 min of moderate-intensity aerobic activity a week because of chronic conditions, they should be as physically active as their abilities and conditions allow.

Moderate-intensity physical activity: Aerobic activity that increases a person's heart rate and breathing to some extent. On a scale relative to a person's capacity, moderate-intensity activity is usually a 5 or 6 on a 0 to 10 scale. Brisk walking, dancing, swimming, or bicycling on a level terrain are examples. Vigorous-intensity physical activity: Aerobic activity that greatly increases a person's heart rate and breathing. On a scale relative to a person's capacity, vigorous-intensity activity is usually a 7 or 8 on a 0 to 10 scale. Jogging, singles tennis, swimming continuous laps, or bicycling uphill are examples. Muscle-strengthening activity: Physical activity, including exercise that increases skeletal muscle strength, power, endurance, and mass. It includes strength training, resistance training, and muscular strength and endurance exercises. Bone-strengthening activity: Physical activity that produces an impact or tension force on bones, which promotes bone growth and strength. Running, jumping rope, and lifting weights are examples.

Source: Adapted from U.S. Department of Health and Human Services. *Physical Activity Guidelines for Americans, 2nd edition*. Washington, DC: U.S. Department of Health and Human Services; 2018. Available at https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf.

Patients with persistent complaints of poor sleep quality or excessive daytime somnolence or with witnessed apneic spells may benefit from screening for sleep disorders, prior to consideration of a formal sleep study. A number of clinical tools have been developed to screen for sleep apnea, including the Epworth Sleepiness Scale, the STOP (snoring, tiredness, observed apnea, high blood pressure) Questionnaire, and the STOP-Bang Questionnaire (STOP plus assessment of BMI, age, neck circumference, and gender), among others. The U.S. Preventive Services Task Force found that current evidence is insufficient to assess the balance of benefits and harms of screening for OSA in asymptomatic adults owing to a lack of validation data in primary care settings. Nonetheless, the high prevalence and significant health consequences of sleep apnea suggest that clinicians should be alert for its potential presence, particularly in patients who are obese with symptoms of excessive daytime somnolence or witnessed apnea episodes. Other sleep disorders, such as restless leg syndrome, may be identified with simple history.

Weight Management Overweight and obesity are prevalent in epidemic proportions in the United States and other industrialized nations (**Chaps. 401 and 402**). Since 1985, the prevalence of obesity in the United States has increased from ~10 to ~35%, and the prevalence of overweight is now ~40%. Overweight and obesity disproportionately affect individuals in lower socioeconomic strata and in many underserved minority populations, including black Americans, Latino Americans, and American Indians. In all race/ethnic groups, both overweight and obesity are associated with adverse health consequences, including diabetes, certain cancers, cardiovascular diseases, and degenerative joint disease. Eating disorders such as anorexia and bulimia are much less common but pose major health consequences

for affected patients and should be suspected particularly in younger women with history of rapid weight shifts or underweight status.

Weight loss is one of the most difficult preventive interventions to achieve and sustain over time. However, several key factors can assist the patient and clinician, and early referral to a dietitian can be very helpful. The first therapeutic goal is to aim for weight stabilization. Many of the risks of overweight and obesity are driven more strongly by continued weight gain, rather than overweight/obese status per se. Working with the patient to find initial strategies for weight maintenance can be a successful initial step with success for many patients. For those who can progress to considering weight loss, it is critical to help the patient understand that there is no standard solution. Experimentation and documentation are key. Tools to assist patients can include food and weight logs, activity logs, and smart phone apps. Some patients respond best to structured approaches such as intermittent fasting regimens or commercial dietary programs where meals are provided. Any of these approaches can be tried with or without social group supports.

The key construct for weight loss is, of course, negative calorie balance. This is achieved through a combination of reduced caloric intake and increased physical activity. Patients may already understand, from prior weight loss attempts, what combination works best for them to achieve this. Some patients find that they cannot lose weight without increasing their exercise. For many, reduction of caloric intake is most efficient. Encouraging the patient to find what works for them is most important. The same principle holds for dietary content. Well-done feeding studies indicate that weight loss is dependent far more on the reduction of caloric intake than on the relative composition of fat, protein, and carbohydrate in the diet. There may be other medical reasons

to choose one approach over another, but if not, encouraging the patient to pick one approach and document the results is an important start. Once weight loss is achieved, increase in activity is often required for its successful maintenance.

Tobacco Cessation (see Chap. 454) Escaping nicotine dependence is another major, but critical, challenge to prevention and wellness efforts. The addictive effects of nicotine have been well documented, with effects that can last for years after successful cessation. Assessing a patient's past history of cessation attempts and current readiness for change are key first steps in forging a successful approach. Frequent follow-up and reinforcement, as well as use of nicotine replacement therapy and other cessation-promoting medications, are additional critical elements. Recidivism is the rule, and patients should expect to resume smoking and attempt again as they journey to tobacco cessation. Electronic cigarettes have some evidence for benefit in adult smoking cessation, but their potential for use by adolescents and young adults who are not smokers represents a major public health threat for a new generation of nicotine addiction, with unknown health consequences as a result of the high doses of nicotine delivered to developing organs, including the brain. Vaping of other substances, often in association with flavoring compounds, has also been associated with pulmonary and cardiovascular damage and should be actively discouraged.

VACCINATION (CHAP. 123)

One of the major advances in public health that has contributed to increases in health and longevity worldwide is the development of safe and effective vaccinations against endemic and epidemic infectious diseases. Patients should be counseled regarding age-appropriate vaccinations for their children and for themselves. Some individuals may be reluctant to receive a vaccination; in these cases, listening to the patient's concerns is important, followed by explanation of the benefits to the individual, their family, and their community and review of the low risk for potential harms. It is true to say that no current vaccines are ever worse than the disease they prevent, although side effects may occur rarely. Thorough knowledge of the data on side effect rates and of efficacy will aid the clinician in helping the patient make a fully informed decision.

MENTAL HEALTH AND ADDICTION

Assessment for depression and cognitive impairment is important to address when patients exhibit symptoms or they or their family members express concerns. Both of these common conditions play a major role in reducing quality of life and are high on patients' lists of concerns, even if not clearly expressed. Screening tools for depression are reviewed in *Chap. 452*. Cognitive function decline with aging or comorbid illness, including depression, should be anticipated. Assessment tools such as the General Practitioner Assessment of Cognition or the Mini-CogTM test are widely available and effective rapid assessment tools.

Alcohol and Opioids (see Chaps. 453 and 456) Alcohol dependence and abuse are common and underdiagnosed. Rapid screening tools have proven efficacy for identifying patients with alcohol problems. In a systematic review, the CAGE (cut down, annoyed, guilty, eye opener) questionnaire was most effective at identifying alcohol abuse and dependence, with reasonable sensitivity and high specificity. The present opioid epidemic in the United States presents a new and substantial public health challenge given the high potential for dependency and abuse of these drugs. Rapid screening tools are available to assist clinicians in screening for opioid dependence.

ACCIDENTS AND SUICIDE

Regular assessment of patient safety through simple questions about seat belt use, domestic violence, and gun safety in the home continues to be an important part of health promotion and wellness. Long-standing recommendations for assessment of suicidal ideation among patients with depression or a history of suicide attempts also continue to be relevant.

APPROACH TO THE PATIENT

In the context of a clinical visit focused on health assessment, health promotion, and prevention, the basic skills of history-taking are of paramount importance. Much of the evaluation, counseling, and management that focus on health promotion and prevention also require engagement and buy-in from the patient in order to assist with recognition of contributing behaviors and to promote adherence to therapeutic plans. Therefore, in addition to standard history-taking, additional skills such as motivational interviewing and eliciting patient commitments and contracting may prove of significant value. The availability of additional tools to assist with screening, monitoring, and chronic management, both online and through wearable devices and mobile health technologies, is rapidly expanding, with uncertain implications for the future. Major research gaps exist in our understanding of how best to employ these newer technologies to improve health outcomes. Concepts of behavioral economics are being explored to better understand the psychology of decision-making and incentives as a means to improve lifestyle choices and adherence to treatment plans (*Chap. 481*).

The limited time available to clinicians and patients during a wellness visit or periodic health examination (not driven by specific patient issues) makes it important to prioritize assessment and counseling for factors that affect longevity, health span, and quality of life over approaches that may have low yield, such as the annual comprehensive physical examination in an asymptomatic patient. Setting clear expectations for the content of a wellness visit may be a first step, and scheduling follow-up visits for findings or to continue indicated counseling are important steps to achieving better health outcomes.

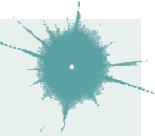
FURTHER READING

- BOULWARE LE et al: Systematic review: The value of the periodic health evaluation. *Ann Intern Med* 146:289, 2007.
- DIETARY GUIDELINES FOR AMERICANS, 2020–2025. Washington, DC: U.S. Department of Agriculture and U.S. Department of Health and Human Services; 2020. Available at https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf.
- IRISH LA et al: The role of sleep hygiene in promoting public health: A review of empirical evidence. *Sleep Med Rev* 22:23, 2015.
- KROGSBOLL LT et al: General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. *BMJ* 345:e7191, 2012.
- U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES: *Physical Activity Guidelines for Americans*, 2nd ed. Washington, DC: U.S. Department of Health and Human Services; 2018. Available at https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf.
- U.S. PREVENTIVE SERVICES TASK FORCE webpage. Available at <https://www.uspreventiveservicestaskforce.org/uspstf/>.

3

Vaccine Opposition and Hesitancy

Julie A. Bettinger, Hana Mitchell



Vaccines have been recognized as one of the top public health achievements of the twentieth century. Dramatic declines in the morbidity and mortality of vaccine-preventable diseases have been observed, and the contribution of vaccines to the elimination, control, and prevention of infectious disease cannot be overstated. However, opposition and hesitancy to vaccines exist and are not new. Vaccine hesitancy has existed

since Edward Jenner introduced the first vaccine against smallpox in the eighteenth century. So why did the World Health Organization rank these attitudes as one of the ten greatest threats to public health in 2019? Are current opposition and hesitancy any different from what has been seen before? Many sociologists, public health experts, and health care providers (HCPs) argue yes. Recent social and cultural trends, combined with new communication formats, have converged to create a particularly potent form of hesitancy and what some have labeled a crisis of confidence. This crisis manifests as a lack of trust in specific vaccines, vaccine programs, researchers, HCPs, the health care system, pharmaceutical companies, academics, policymakers, governments, and authority in general. (**See “Focus: COVID-19 Vaccine Hesitancy,” below.**)

The roots of modern vaccine hesitancy and opposition—defined as delay or rejection of vaccines in spite of availability—vary depending on the place and the population. For some individuals and communities, pseudoscience and false claims about the safety of existing vaccines (e.g., an unsupported link between measles vaccine and autism) have driven fears, increased hesitancy, and decreased acceptance. For others, real safety events, such as the association of narcolepsy with a specific pandemic influenza vaccine (Pandemrix), have justified concerns. In a few locations (e.g., Ukraine, Pakistan), vaccine hesitancy is the result of failed health systems or even state failures. Finally, for some groups, including some fundamentalist religious groups and alternative-culture communities, vaccine hesitancy and opposition reflect exclusion from and rejection of mainstream society and allopathic health care and manifest as a deep distrust of these institutions and their HCPs. Although the genesis of modern vaccine hesitancy is multifactorial, its outcomes are uniform: a decrease in vaccine demand and uptake, a decrease in coverage by childhood and adult vaccines, and an increase in vaccine-preventable diseases, outbreaks, and epidemics of disease. Addressing this crisis and moving people from vaccine hesitancy and refusal to acceptance and active demand require intervention at multiple levels: the individual, the health system (including public health), and the state.

This chapter will define vaccine hesitancy and briefly describe its determinants and effects in North America (the United States and Canada). Physicians and other HCPs are well positioned to address the

crisis of confidence many patients feel toward HCPs and the health care system. Studies demonstrate that an unambiguous, strong recommendation by trusted HCPs is most often the reason that patients, including those who are vaccine hesitant, choose to vaccinate. Strategies for counseling vaccine-hesitant and vaccine-resistant patients will be presented and examples of strong vaccine recommendations provided. Presenting strategies to increase vaccine demand at a system and policy level is beyond the scope of this chapter. While some physicians may have roles that allow them to act at this level, all physicians can act and influence their individual patients. Strategies to create active vaccine demand at the individual level alone will not solve vaccine hesitancy, but vaccine hesitancy cannot be addressed without these efforts. **For further discussion of immunization principles and vaccine use, see Chap. 123.**

VACCINE COVERAGE AND OUTBREAKS

The epidemiologic data from measles outbreaks over the past 10 years provide an interesting illustration of the effects of vaccine opposition and hesitancy. **For further discussion of measles, see Chap. 205.**

North America *Herd immunity* occurs when enough individuals in a population become immune to an infectious disease, usually through vaccination, that transmission of the infection stops. The level of immunity (or level of vaccine coverage) required to confer herd immunity varies with the specific infectious disease. Because measles is a highly contagious virus, a coverage rate of 93–95% must be achieved for vaccination to confer herd immunity and interrupt measles transmission. National coverage estimates place one-dose measles vaccine coverage rates in 2-year-old children at 92% in the United States and 88% in Canada. In spite of these relatively high levels of coverage in young children, numerous measles outbreaks have occurred in both countries since 2010 (**Table 3-1**).

The vast majority (>80%) of measles cases described in Table 3-1 occurred in under- or completely unvaccinated individuals. Of note, many of these outbreaks highlight pockets of significantly under- or unvaccinated individuals that are not apparent in national vaccine coverage statistics. Moreover, many of the outbreaks listed in Table 3-1 were ignited by unvaccinated returned travelers from areas with existing

TABLE 3-1 Measles Outbreaks in North America

YEAR/PLACE	NO. OF CASES	REASON
2010/Canada	70	An infected traveler to the 2010 Winter Olympics transmitted infection to an under- and unvaccinated local population in British Columbia.
2011/Canada	776	Disease was imported from France by an unvaccinated returned traveler to Quebec. The outbreak spread in a nonvaccinating religious community and outside that community. A majority of cases occurred in under- and unvaccinated persons.
2011/United States	118	Of 118 cases, 46 were in returned travelers from Europe and Asia/Pacific regions; 105 cases (89%) occurred in unvaccinated persons.
2013/United States	58	Disease was imported by a returned unvaccinated traveler from Europe. The outbreak spread in a nonvaccinating religious community in New York.
2014/Canada	433	Disease was imported from the Netherlands. The outbreak spread in a nonvaccinating religious community in British Columbia.
2014/United States	383	The outbreak occurred in nonvaccinating religious communities in Ohio.
2015/United States	147	A multistate/multicountry outbreak was linked to Disneyland amusement park. More than 80% of cases occurred in unvaccinated persons.
2015/Canada	159	Disease was imported from the United States (part of the Disneyland outbreak) by an unvaccinated traveler. The outbreak spread in a nonvaccinating religious community in Quebec.
2017/United States	75	The outbreak occurred in an under-vaccinated community in Minnesota; 95% of patients were unvaccinated.
2018/United States	375	Disease was imported by returned unvaccinated travelers from Israel. The outbreak spread in nonvaccinating religious communities in New York and New Jersey.
2019/Canada	31	Disease was imported from Vietnam by a returned traveler to British Columbia. The outbreak spread throughout local area schools in under- and unvaccinated persons and resulted in a province-wide measles mass immunization campaign for schoolchildren.
2019/United States	1282	Outbreaks occurred in 10 states; 73% of cases (~935) were linked to outbreaks in nonvaccinating religious communities in New York.

Source: Centers for Disease Control and Prevention and Public Health Agency of Canada.

outbreaks or epidemics, who spread disease into an unvaccinated or under-vaccinated community. Many of the outbreaks were contained within the nonvaccinating community, but several spread to other under-vaccinated communities geographically contiguous with the outbreak community. More concerning still are the cases and outbreaks originating in communities that had not previously been identified as nonvaccinating. These cases likely highlight pockets of unvaccinated individuals who object for cultural rather than religious reasons. In the past, these nonvaccinating individuals did not exist in large enough clusters to sustain the spread of measles. Of further concern is the number of individuals included in outbreak statistics who have had one or sometimes even two doses of vaccine and who were thought to be protected but who still end up with the disease. The assumption is that one or two doses provide full disease immunity, but this is not always true. Often, individual level characteristics (age, immune compromise, etc.) affect the individual's response to the vaccine and their level of protection. In other instances, vaccine protection can wane over time, thus leaving fully immunized individuals susceptible to infection. In fact, when herd immunity breaks (i.e., the level of immunity in a community becomes too low to prevent transmission of disease), the occurrence of cases even in fully immunized persons is seen, as reflected in outbreak statistics. As a result of decreased vaccination rates and the resulting disruption of herd immunity, these individuals may become more identifiable as non-immune.

Outside North America Although overall coverage rates may still be high in North America, they are lower in other parts of the world. In Samoa, for example, measles–mumps–rubella (MMR) vaccine coverage before a recent outbreak was 31%; in the Philippines, it was 67%. Twenty years ago, vaccine coverage was sufficiently high in some parts of the world, including Europe, that an unvaccinated traveler from a nonvaccinating community to most regions would have been protected by herd immunity at their destinations. Today that is not the case: such travelers are likely to become infected in a country with active measles transmission and return home to spread the infection into their communities and possibly beyond. Thus active measles transmission, whether at home or abroad, places individuals who rely on herd immunity (e.g., immunocompromised persons and young infants) at increased risk.

FACTORS IN VACCINE HESITANCY

Vaccination coverage rates provide an estimate of the proportion of children or adults in the population who have been vaccinated, but they do not indicate the proportion of individuals who are vaccine hesitant. An individual may be fully vaccinated but still be hesitant about the safety and effectiveness of vaccines, or an individual may be unvaccinated as a result of access issues but may not be hesitant. Therefore, in attempts to understand a patient's lack of vaccination, it is important to distinguish persons who are hesitant and refuse vaccines from those who need assistance to access the health care system and successfully complete vaccination. To this end, an understanding of vaccine hesitancy and its determinants is needed.

Vaccine hesitancy and opposition are defined by the World Health Organization's SAGE Working Group on Vaccine Hesitancy as a "delay in acceptance or refusal of vaccines despite availability of vaccination services." The SAGE group describes vaccine hesitancy as "complex and context specific, varying across time, place, and vaccines."

It is useful to frame vaccine acceptance as a continuum pyramid, with active demand for all vaccines representing the largest group at the bottom of the pyramid and outright refusal of all vaccines depicted in the smallest group at the top. In the middle lies vaccine hesitancy, in which the degree of vaccine demand and acceptance varies. Fortunately, for disease control efforts, most individuals fall within the active-demand category or, if they are hesitant, still accept all vaccines. Hesitancy can be influenced by complacency, convenience, and confidence (Fig. 3-1).

Complacency is self-satisfaction when accompanied by a lack of awareness for real dangers or deficiencies. Complacency exists in communities and individuals when the perceived risks of vaccine-preventable diseases are low and vaccination is not deemed a necessary preventive action. This attitude can apply to vaccination in general or to specific vaccines, such as influenza vaccines. Actual or perceived vaccine efficacy and effectiveness contribute to complacency. Patients who are complacent about vaccine-preventable diseases prioritize other lifestyle or health factors over vaccination. These individuals can be influenced toward vaccination by a strong recommendation from a trusted HCP or a local influenza outbreak. They can be influenced away from vaccination by a vaccine scare or misinformation on social

Characteristics

- Strong distrust of health system/pharmaceutical industry/government
- Strong-willed and committed against vaccines
- Negative or traumatic experiences with HCPs and health system
- May use natural approach to health/alternative HCPs
- May have strong religious/moral considerations for refusal
- May cluster in communities (geographic and online)
- Vaccination is very unlikely; alternative strategies to protect individual and community must be discussed.

- Questions safety and necessity of vaccines
- Actively seeks information from many sources
- Has conflicting feelings on whom to trust
- Social norm is not vaccinating.
- May have had negative or traumatic experience with health system
- Vaccination may not occur; a strong trust relationship with HCP and many visits and conversations are required.

- Focused on vaccine risks
- Conversation with trusted HCP strongly influential
- Trusts HCPs
- Actively seeking information and wants to verify it
- Wants advice specific for their child
- Confused by conflicting information
- Social norm is vaccinating, but individual may feel conflicted by this norm.
- Vaccination requires longer conversation and may require multiple visits.

- Focused toward vaccine risk
- **Complacency:** low perceived benefits of vaccination
- Can move up or down continuum as a result of various influences (HCP recommendation, vaccine scare, outbreak)
- Trusts HCPs and health system
- **Convenience:** need few barriers to vaccination
- Vaccination requires longer conversation but likely can be performed at same visit; potential exists to move to active demand.

- **Confidence**
- Considers vaccines important
- Considers vaccines safe
- Trusts HCP/vaccines/health system
- Social norm is vaccinating
- Very short conversation with HCP about vaccination, in which HCP should address any questions to maintain active-demand status

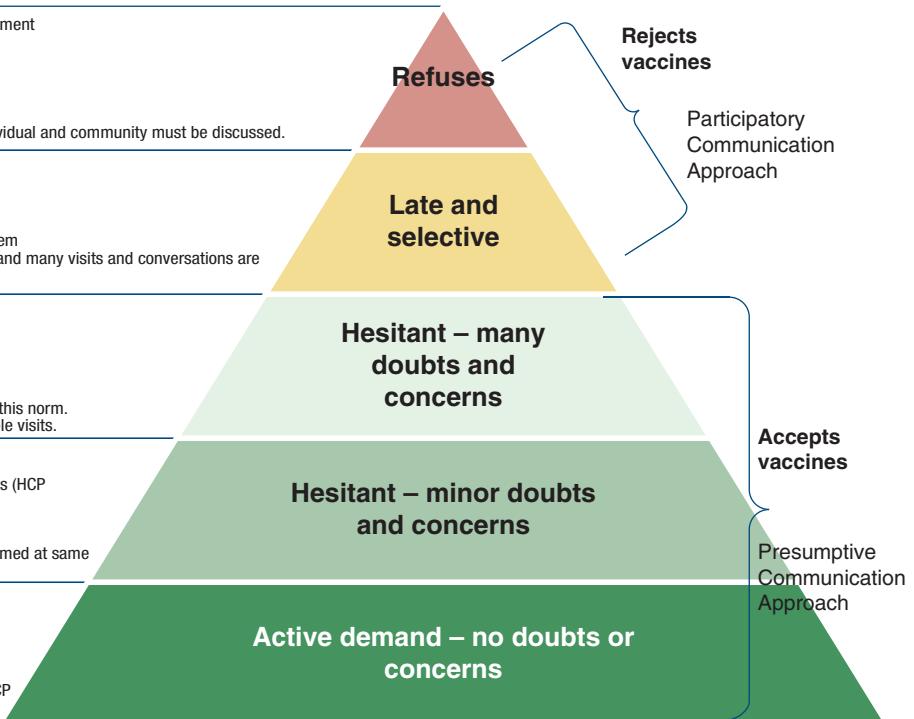


FIGURE 3-1 Vaccine acceptance continuum. HCPs, health care providers. (Adapted from J Leask et al: BMC Pediatrics 12:154, 2012; AL Benin et al: Pediatrics 117:1532, 2006; and E Dubé, NE MacDonald: The Vaccine Book, 2016, pp. 507-528.)

media. Finally, the real or perceived ability of patients to take the action required for vaccination (i.e., self-efficacy) influences the role complacency plays in hesitancy and willingness to seek vaccination.

Convenience is determined by the degree to which conversations about vaccination and other services can be provided in culturally safe contexts that are convenient and comfortable for the individual. Clearly, convenience varies by community, health clinic, and even patient. Persons who are criticized or scolded for not vaccinating themselves or their children may not feel comfortable or safe accessing health services. Factors such as affordability, geographic accessibility, language, and health literacy are important considerations when evaluating the convenience of existing clinical care. Any of these factors can affect vaccine acceptance and can push a patient who has some hesitancy toward vaccinating or not vaccinating.

Confidence is based on trust in the safety and efficacy of vaccines, in the health care system that delivers vaccines (including HCPs), and in the policymakers or governments who decide which vaccines are needed and used. A continual erosion of confidence around vaccination, health systems, and governments drives today's hesitancy and has been amplified by larger social and cultural trends in medicine, parenting, and information availability.

SOCIAL AND CULTURAL TRENDS

Individualized Health Care Over the past 30 years, the focus of medicine and health care has shifted to patient-oriented, individualized care, with an increasing emphasis on treatment and prevention options tailored to the individual patient. In vaccination programs, this shift has manifested as requests for individualized vaccine recommendations and customized immunization schedules. The increasing personalization of medicine, while positive overall, has forced public health away from a focus on the community and its common good and has created tension between individual rights and community health.

Parenting Trends The desire for an individualized approach to medicine and vaccination reflects broader cultural trends concerning individual risk management: accordingly, the individual is to blame for bad outcomes, and public institutions cannot be trusted to manage technological (i.e., vaccine-related) risks. This viewpoint is directly linked with cultural shifts in parenting and social norms defining what it means to be a "good parent." The image of a good parent has been reframed to refer to someone whom several investigators have described as "a critical consumer of health services and products, accounting for their own individual situation as they see it with little regard for the implications of their decision on other children." The archetypical good parent no longer unquestioningly trusts HCPs and other authorities and experts. According to this social norm, "good parents" should seek individual medical advice that is tailored for their child and specific to that child's needs. While in essence not a bad thing, this norm can conflict directly with public health vaccine recommendations and schedules that are organized to maximize community health and to facilitate efficient provision of care at a community level.

Traditional Media Newspapers, radio, and television have been criticized for their coverage of vaccines and in particular their coverage of the alleged link between MMR vaccine and autism. By offering equal coverage throughout the early to mid-2000s for both the scientific evidence and unproven claims of MMR vaccine harms, traditional media outlets provided a forum and a megaphone for the spread of pseudoscience. Equal coverage leads to false equivalencies. Celebrity advocates further amplified the message via this channel. The boost that traditional media provided to active vaccine resistance and, less directly, to vaccine hesitancy has not been adequately measured but must be considered in any discussion of vaccine hesitancy. After headlines about multiple outbreaks of measles and other vaccine-preventable diseases and continued direct criticism of the equal-coverage approach, some traditional media now reject and attempt to discredit pseudoscience. The effect this stance will have on increasing vaccine confidence is unknown.

The Internet and Social Media Approximately 90% of Americans and 91% of Canadians use the Internet, and 80% of Americans and 60% of Canadians have a social network profile. Widespread access to social media can be empowering, but it is also problematic. The Internet and social media require users to select their information sources, creating an environment described as an "echo chamber" in which individuals choose information sources harboring beliefs or opinions similar to their own and thereby reinforcing their existing views. This situation has created a new platform for further spread of vaccine *misinformation* (inaccuracies due to error) and *disinformation* (deliberate lies) and has provided a forum for vaccine-resistant individuals, including celebrities, to organize and raise funds to support their efforts. The harmful effects of Internet and social media use on vaccine hesitancy have been well documented. Vaccine hesitancy increases for parents who seek their information from the Internet. Unfortunately, public health and health care institutions have been slow to adapt to this new communication medium and to recognize its influence and impact. In this medium, personal stories and anecdotes are now viewed as data and disproportionately influence vaccine decision-making, while traditional, more authoritative, fact-based information sources are deemphasized. Centralized monitoring by jurisdiction of vaccine misinformation and disinformation, with summaries of the relevant discourses and rebuttals provided to HCPs, has been proposed as a potential way to counter the influence of social media on vaccine hesitancy. While such strategies have been applied in single jurisdictions and appear to have had some success, their applicability to a broader context is unknown. Moreover, the resources for such a coordinated response have not been made available, and individual HCPs have been left to counter popular, shifting, viral communications on their own, patient by patient.

As with traditional media, the social media landscape appears to be shifting. In 2019, the proliferation of anti-vaccination information combined with measles outbreaks in North America and increasing pressure from health leaders led large social media companies (Facebook, Instagram, Pinterest) to deemphasize anti-vaccination information by removing relevant advertisements and recommendations and decreasing their prominence in search results. While it is too soon to determine the effects of these measures, critics are skeptical that they will have the intended result of reducing vaccine misinformation and disinformation. Early evidence shows that misleading content is still widely available, with anti-vaccine advertisements now using the term "vaccine choice" to avoid censorship. More disturbingly, public health advertisements in support of vaccination have been included in social bans and removed from social media sites.

In a more grassroots effort, providers and vaccine supporters have united on social media to provide online support and evidence-based facts to providers and others who support vaccines when they are attacked digitally by anti-vaccine supporters. For example, Shots Heard Round the World (www.shotsheard.com) is an effort led by two U.S. pediatricians to provide advice and support for HCPs who speak out about the importance of vaccines. Such efforts harness the power of social media in ways similar to those used by vaccine opponents and may prove successful in combating vaccine hesitancy.

Given these social and cultural trends, no one should be surprised when individuals now question vaccination, express confusion about conflicting information and information sources, and feel unsure whom to trust. Their broader social context is telling them they should question everything and trust no one. This message is reinforced via misinformation and disinformation on social media. Recent vaccine-preventable disease outbreaks illustrate that effective engagement with individuals cannot be accomplished through one-way, top-down information provision (which still is often the *de facto* choice for health system communication), but rather requires a dialogue that takes into account the social processes surrounding individual vaccination decisions. It is at the interface between the individual and the health system in which conversations between HCPs and their patients can have the greatest impact. It is critical for all HCPs to discuss vaccines and provide strong vaccine recommendations—including HCPs who do not administer vaccines but who have established trust with their patients.

APPROACH TO THE PATIENT

An ideal vaccine-hesitancy intervention would result in full compliance with vaccination, the patient's satisfaction with the health care encounter, and sustained trust in the HCP's recommendations. On a programmatic level, vaccine-hesitancy interventions should be multicomponent, dialogue based, and tailored to specific under-vaccinated populations.

Communicating with vaccine-hesitant individuals can be challenging and time-consuming. HCPs may feel that vaccine-hesitant patients cast doubt on their personal and professional integrity, their authority as medical experts, and their competence as communicators. Some HCPs may be reluctant to initiate conversations about vaccination because of concerns that discussing a sensitive topic may compromise their clinical rapport with their patients. Other HCPs may believe that they have not received sufficient training to confidently recommend vaccines and answer questions. Discussing vaccines with hesitant patients, while not always easy, provides an opportunity to honor the principles of patient-centered care by demonstrating an interest in patients' opinions, engaging in dialogue, and ideally increasing patients' confidence in vaccine recommendations.

FACTORS IN EFFECTIVE VACCINE RECOMMENDATIONS

Vaccine recommendations ideally should be made within an established, trusting patient-provider relationship in which patients are comfortable asking questions and voicing concerns, even if their views on vaccines contradict the HCP's recommendations. Recommending vaccines requires both provision of information and effective communication. There is no single "best practice" for how providers should approach recommending vaccines to vaccine-hesitant individuals. In general, all vaccine recommendations should be (1) strong, making it clear that the provider supports and recommends vaccination; (2) tailored, acknowledging the vaccine attitudes and potential concerns of individual patients; (3) transparent and accurate, highlighting the benefits of vaccines while also communicating the risks; (4) supported by trustworthy information resources that patients can access and review after the clinical encounter; and (5) revisited, with repetition and reinforcement during follow-up health care encounters.

Strength of the Recommendation HCPs should make it explicit (in the absence of medical contraindications) that vaccination based on the recommended schedule is the best option. While HCPs should take time to elicit patients' questions and address concerns, the recommendation for vaccination should be made in clear and unambiguous terms.

Tailored Communication Vaccine hesitancy occurs on a continuum (Fig. 3-1). Therefore, it is helpful for HCPs to have some understanding of their patients' attitudes toward vaccination at the start of the health care appointment. Unfortunately, vaccine-hesitancy surveys for use as part of vaccine consultation visits have not been validated on a large scale. However, the following are some examples of questions that can be asked, depending on the setting. (1) Did you have a chance to review the vaccine leaflet we provided? Did you have any questions about it? (2) Have you ever been reluctant or hesitant about getting a vaccination for yourself or your child? If so, what were the reasons? (3) Are there other pressures in your life that prevent you from getting yourself or your child immunized on time? (4) Whom/what resources do you trust the most for information about vaccines? Whom/what resources do you trust the least?

Communication style and content for patients in the active-demand category for vaccination will be different from those for individuals who are hesitant, late and selective, or strongly inclined to refuse vaccines. Two communication styles have been proposed for vaccine recommendations. Evidence shows that a *presumptive/directive* approach ("Your child is due for MMR vaccination.") results in higher rates of vaccine uptake than a *participatory/guiding* approach ("What are your thoughts about the MMR vaccine?").

However, adopting a strictly presumptive/directive approach may alienate some patients, especially those who are higher up on the hesitancy pyramid and who may feel that they are being pressured into vaccination before their concerns have been heard and addressed. Adopting a participatory/guiding approach and clarifying receptivity to vaccines may be more suitable for hesitant individuals with many doubts and concerns, persons with a late or selective attitude, and those who are strongly inclined to refuse vaccines. In addition, a participatory/guiding approach provides an opportunity for ongoing clinical rapport and dialogue between unvaccinated or under-vaccinated patients and their HCPs, even when it does not result in immediate vaccine uptake. Regardless of which approach is used, a strong vaccine recommendation should be made at each encounter.

Transparency and Accuracy Vaccine recommendations should be transparent, should include accurate information about both the benefits and the risks of the vaccine, and should emphasize why the benefits outweigh the risks. For example, when evidence supports an association between a vaccine and an adverse event, the occurrence of the adverse event is often very rare and the event quickly resolves (Chap. 123). U.S. Federal law (under the National Childhood Vaccine Injury Act) requires HCPs to provide a copy of the current Vaccine Information Statement from the Centers for Disease Control and Prevention (CDC), which describes both benefits and risks of vaccines to an adult patient or to a child's parent/legal representative before vaccination.

CDC Vaccine Information Statements should not replace a discussion with the HCP. Depending on the provider and the patient, a description of benefits and risks may include words and numbers, graphics, and personal anecdotes (e.g., why the provider vaccinates his or her own children). Personal anecdotes are powerful, and many hesitant patients seek and are influenced by them.

A discussion of benefits and risks provides an opportunity to address specific misconceptions about a particular vaccine or about vaccines overall. For example, patients may be concerned about adverse events following vaccination that are not supported by evidence, such as autism following MMR vaccination or myocardial infarction following influenza vaccination in the elderly.

Most adults—even those whose children are fully immunized—still have questions, misconceptions, or concerns about vaccines that should be addressed. A risk/benefit discussion allows HCPs to describe the vaccine safety monitoring systems in place. Providers should emphasize that vaccines are developed and approved through a highly regulated process that includes licensure clinical trials, review and approval by designated regulatory authorities (e.g., the U.S. Food and Drug Administration, Health Canada), strict manufacturing regulations, and ongoing postmarketing safety surveillance.

Support from Accessible Information Sources All vaccine recommendations should be supported by additional information sources patients can assess after the health care encounter. HCPs play an important role as information intermediaries for their patients. They can navigate information (and misinformation) about vaccines and direct patients toward reliable, appropriate resources. HCPs should consider what resources will be suitable for a patient or patient population. Vaccine information resources are available in different media formats and use a combination of images and text to communicate the information to various audiences. See "Further Reading," below, for suggestions or refer to resources provided by local health authorities.

Revisiting and Reinforcement of Vaccine Recommendations All health care encounters offer an opportunity to revisit and reinforce vaccine recommendations. Vaccine-hesitant individuals who do not accept vaccines but are willing to review information should be offered a follow-up appointment to reinforce previously made recommendations and address further questions. Vaccine-hesitant

patients who accept vaccines should be seen at a follow-up appointment to confirm and document vaccine receipt (if vaccine is not given at the point of care), ascertain whether the vaccine was well tolerated, and reinforce the message about vaccine safety and effectiveness. Patients who actively demand vaccines usually do not require much follow-up other than to confirm and document the receipt of vaccine (if it is not given at the point of care) and to address additional questions or concerns arising subsequent to vaccination. Often this follow-up can be covered without an office visit.

WHAT TO SAY TO VACCINE-HESITANT PATIENTS

Engaging vaccine-hesitant individuals requires confidence, knowledge, skills, time, and creativity to tailor the approach to each individual patient. Examples for each part of the vaccine recommendation are listed in **Table 3-2**.

OTHER CONSIDERATIONS DURING CLINICAL ENCOUNTERS

Missed Opportunities The World Health Organization defines a missed opportunity for vaccination as “any contact with health services by an individual (child or person of any age) who is eligible

for vaccination (e.g., unvaccinated or partially vaccinated and free of contraindications to vaccination), which does not result in the person receiving one or more of the vaccine doses for which he or she is eligible.” HCPs who do not offer point-of-care vaccination frequently miss the opportunity to recommend vaccines to their patients. Missed opportunities for recommending and providing vaccines during routine health care encounters contribute to under-vaccination. Studies show that up to 45% of under-vaccinated children could be up to date with all age-appropriate vaccines and up to 90% of female adolescents could be up to date with human papillomavirus (HPV) vaccination if all opportunities to vaccinate were taken.

Vaccine counseling and vaccination should be incorporated into clinical care for individuals of all ages, not just young children. Because many adolescents and adults do not have regular health care follow-up, providers need to take advantage of every health care encounter to recommend and provide vaccines. For example, a visit to an emergency department, a routine follow-up visit at a diabetes clinic, or a visit planning for elective orthopedic surgery offer opportunities to inquire about the patient’s vaccination status and to recommend vaccines.

HCPs should make preemptive vaccine recommendations (e.g., initiating discussions about infant vaccines during pregnancy, informing parents about HPV vaccine before their child becomes eligible). Such advance discussions may be especially helpful in identifying

TABLE 3-2 Sample Vaccine Conversations

STRONG VACCINE RECOMMENDATION

“We are headed into the flu season. Getting flu vaccine not only protects you, but it helps protect other people around you who can get very sick from flu. I strongly recommend you get your flu shot. Do you know where to get it?”

“You will be turning 50 next year. This means you will be eligible for a vaccine that prevents shingles, and I strongly recommend you receive it. Have you heard about this vaccine before? Can I answer your questions about it?”

“I know you are not comfortable getting vaccinated today. I do want to make it clear that I recommend vaccines because I am convinced they are the best way to protect you from some serious diseases. Is there something that would lead you to think about getting vaccinated in the future?”

TAILORED COMMUNICATION

“I recommend that children and adults stay up to date on recommended vaccines. I see from your vaccine record that you’ve had your childhood vaccines, but you haven’t gotten any adult vaccines. I wanted to clarify whether this is because you decided not to get vaccines or something else prevented you from getting vaccinated.”

“I understand that you are here for your pneumococcal vaccine. This is the best way to protect yourself and those around you from pneumonia. Do you have any questions before I give you the vaccine?”

“I understand you have some concerns about vaccines. What are you most concerned about? Would you like me to explain why I recommend giving your child these vaccines?”

TRANSPARENCY AND ACCURACY

“Serious side effects can develop after MMR vaccination but are very rare. On average, 3 out of 10,000 children who get MMR vaccine will have a febrile seizure/convulsion in the days after vaccination. Febrile seizures can be frightening, but nearly all children who have a febrile seizure recover very quickly and without any long-term consequences. On the other hand, 1 out of 1000 children who get measles will develop encephalitis (brain inflammation) that not only causes seizures but can also lead to permanent damage.”

“About 10 out of every 10,000 Americans who do not get vaccinated against flu die because of influenza every year, and many more are hospitalized. While flu vaccine does not prevent all cases of influenza, it is the most effective vaccine we have. By getting the vaccine, you also help protect people around you from getting sick.”

“You are correct, aluminum is used in some vaccines to help the body’s immune system respond. However, aluminum is also present in food and drinking water. In fact, the amount of aluminum present in vaccines is similar to or less than what is present in breast milk or infant formulas.”

SUPPORT FROM ACCESSIBLE INFORMATION SOURCES

“Your child and other boys and girls his age will be eligible for the human papillomavirus vaccine this coming school year. Have you heard about this vaccine before? What questions do you have about it? Here’s a list of websites for parents and teenagers that explain what it is about.”

“There’s a lot of information about vaccines on the internet, and a lot of that information is not based on facts. Here is a list of websites that have been reviewed by health care professionals and accurately describe benefits and risks of each vaccine. The information is written in lay language and includes helpful illustrations.”

REVISITING AND REINFORCEMENT OF THE RECOMMENDATION

“During our last visit, we talked about MMR vaccine for your son and some of the concerns you had about potential side effects. Have you had a chance to look at the take-home information I gave you? Was there anything else you would like to ask about? I recommend that we vaccinate your child today.”

“During our last visit, we talked about receiving a pertussis booster during pregnancy and where you can get vaccinated. Have you had a chance to get your pertussis vaccine?”

“I see that you got your vaccines at the public health clinic last week. How did it go? Did you have any questions?”

“It’s possible that the symptoms you experienced after receiving the vaccine were an adverse reaction to the vaccine. I will report this to the health authority. Let’s discuss what we can do next time to prevent symptoms from occurring again.”

Note: Specific vaccine recommendations, vaccine eligibility guidelines, and statistics used to communicate benefits and risks will vary with the health jurisdiction and the country. Several sample statements here are adapted from the Australian National Centre for Immunisation Research and Surveillance website (www.talkingaboutimmunisation.org.au). For patient vaccine information resources, see also the Immunization Action Coalition website for the public developed in partnership with the CDC (vaccineinformation.org).

vaccine-hesitant patients and ensuring that they have enough time to ask questions and make decisions before vaccines are due.

HCPs should ensure that a vaccine recommendation is followed by vaccination. Providers who recommend vaccines but do not vaccinate at the point of care should inform patients where they can be vaccinated. This discussion may include information about public health clinics, travel clinics, and pharmacies or a referral to another provider. HCPs should follow up with their patients at subsequent appointments to confirm that they were vaccinated.

Adverse Events Following Vaccination Although rare, adverse events ([Chap. 123](#)) may influence vaccine acceptance and willingness to be vaccinated in the future. It is important for providers to identify and follow up with all patients who experience an adverse event, regardless of the patients' vaccine attitudes prior to the event. Adverse events following vaccination should be reported to the relevant vaccine monitoring system: the U.S. Vaccine Adverse Event Reporting System or the Canadian Adverse Event Following Immunization Surveillance System.

Addressing Inequities In Vaccine Access Discrepancies in access to health care services create inequitable access to vaccines for children and adults and contribute to under-vaccination. A U.S. study found that socially disadvantaged individuals were more likely than other persons to be under-vaccinated, in part because of a lack of access to health care services. HCPs must recognize that socially disadvantaged individuals and populations are often at greater risk of vaccine-preventable diseases (e.g., as a result of crowded living conditions, limited access to sanitation, poor nutrition, or substance abuse) and also at greater risk of being under-vaccinated because they have limited access to health care services. In addition, specific vaccines may be recommended for some socially disadvantaged populations or communities. For example, in the wake of several outbreaks of hepatitis A among the U.S. homeless population, the CDC now recommends that everyone >1 year of age experiencing homelessness receive hepatitis A vaccine.

Depending on the setting and the patient, some recommended vaccines may not be covered through public funding or private insurance coverage. HCPs should be aware of alternative funding models, such as the Vaccines for Children Program, which provides free vaccines for U.S. children (<19 years of age) with financial barriers to vaccine access. When vaccines are not publicly funded or covered by private insurance and patients perceive that they cannot afford a vaccine, HCPs should not withhold a vaccine recommendation. The risks and benefits of vaccination still need to be communicated, with a strong recommendation, and the patient should be provided the opportunity to decide whether they can afford the vaccine.

Further Communication With Patients Who Refuse Vaccines Fortunately, the proportion of people who completely refuse all vaccines and are not willing to talk to their HCP is small. Nevertheless, in some cases, attempts to initiate discussion and address vaccine refusal may be futile. When possible, HCPs should focus on the common goals of care and preserve the therapeutic relationship. Vaccine refusal should be well documented in the patient's chart. The HCP should continue with tailored communication and be open to future discussions. Vaccine demand and vaccine refusal are rarely static over time. ([See “Focus: COVID-19 Vaccine Hesitancy,” below.](#))

CONCLUSION

In summary, vaccine hesitancy is complex and context specific. It varies with time, place, patient, and vaccine. HCPs are well positioned to address vaccine hesitancy and should develop the skills, knowledge, and confidence to make strong vaccine recommendations to their patients.

FOCUS: COVID-19 VACCINE HESITANCY

As COVID-19 vaccines are used to control SARS-CoV-2, some individuals will have concerns about these vaccines and a proportion of

the population will reject them. While worrisome, hesitancy about COVID-19 vaccines is not unexpected; it mirrors public concerns expressed about past pandemic influenza vaccines and other newly introduced vaccines. It has been established that the newness of any vaccine, be it a pandemic influenza vaccine or a COVID-19 vaccine, raises concern in a large percentage of the population. Politicization of COVID-19 vaccines raises additional issues for some patients.

Past Experience with New Vaccines Past experience with new vaccines, including the H1N1 pandemic influenza vaccine in 2009 and the human papillomavirus vaccine in the early 2000s, provides a guide to topics that need to be addressed with regard to COVID-19 vaccines. While resistance is often framed as uncertainty about a vaccine's "newness," further discussion translates this uncertainty into concern about the new vaccine's safety. This concern encompasses both short- and long-term side effects. Frequent, acute adverse effects can be captured in clinical trial data, whereas worries about rare and long-term side effects can be addressed only by direct evidence after the initiation of a new vaccination program. In addition to queries about the overall safety of the vaccine, HCPs can expect specific questions regarding the safety of individual ingredients included in the vaccine, whether or not these ingredients are new and whether or not relevant safety data are available. Information on the incidence of common or expected health events in an unvaccinated population (i.e., background rates) over a 4-week period is helpful in distinguishing what is normal and expected from a point of concern. Studies that have examined this issue with regard to other vaccines can be used as a basis for presenting background rates of expected events in the context of COVID-19 vaccines for some groups; however, it is important to ensure that more specific background-rate information is available to HCPs with regard to the individual groups being vaccinated. HCPs, public health programs, and vaccine manufacturers can anticipate these questions and should develop answers and information to respond to them.

Specific Concerns about COVID-19 Vaccines While some concerns can be anticipated on the basis of past experience with new vaccines, several characteristics of COVID-19 vaccines require new approaches to adequately address individual concerns, and HCPs need to educate themselves in several specific areas. First, an overwhelming amount of attention has been paid to the speed of development of COVID-19 vaccines, with some jurisdictions even skipping the usual clinical-trial steps in an effort to provide vaccine more rapidly to their populations. This situation directly increases concerns about the "newness" of the vaccine and its safety and, unfortunately, raises questions about the entire vaccine development process. Education is required to explain how a process that normally requires 5–10 years was condensed to this degree. (See Lurie et al [2020] for an excellent explanation of the COVID-19 vaccine development process.) In addition, transparency with regard to clinical trial data is required to enable scientists, HCPs, and consumers to read and understand the development and evaluation processes. The usually shrouded, proprietary development process is unsuitable if the final vaccine product is to garner public trust. Education on existing vaccine-safety monitoring systems also needs to be provided. HCPs must familiarize themselves with the vaccine development process and safety monitoring systems if they are to present this information to their patients.

Second, several newer vaccine platforms that are being used for COVID-19 vaccines (e.g., nucleic acid-based vaccines, viral vector) have not been used in the past. This novelty exacerbates public concern about the unfamiliarity of new vaccines and further heightens misgivings about vaccine safety and the potential for long-term adverse effects. Again, HCPs need to familiarize themselves with the new technology and develop effective messaging for their patients. Public health officials have developed resources to address this issue ([see \[www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html\]\(http://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html\)](http://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html)), but, even in the absence of such resources, HCPs can anticipate questions about the new technology involved and become comfortable explaining it.

Third, clinical trial safety and efficacy data were lacking for all groups initially prioritized to receive the vaccine. For example, long-term-care residents were prioritized for vaccine receipt, but clinical trial data were not available for the range of chronic health conditions that exist in older adults. While observational studies have filled some of these gaps, HCPs need to extrapolate on the basis of available evidence in considering individual patients and must make a recommendation without knowing all the answers.

Fourth, some minority and marginalized communities who have been disproportionately affected by COVID-19 express hesitancy or reject COVID-19 vaccines. For some Black, Indigenous, Latinx, and other communities, COVID-19 hesitancy stems directly from systematic discrimination, racism, and mistreatment in the health care system. Black and Indigenous communities also share a horrific legacy of unethical medical experimentation,¹ which, when combined with current discrimination and overt racism, creates a powerful climate of mistrust in HCPs, the medical system, and science.

Social and Cultural Trends The social and cultural trends already discussed in this chapter—in particular, traditional media, the Internet, and social media—are exerting influence and pressure that did not affect the introduction of older vaccines, even the H1N1 pandemic vaccines. The media attention given to the development of transverse myelitis in one clinical-trial participant following receipt of COVID-19 vaccine is but one example of the intense media scrutiny of the vaccine development process. Unfortunately, in the United States, efforts to control COVID-19, including vaccine development, have become highly politicized. This degree of politicization has not occurred with past vaccines, so HCPs are in uncharted territory in terms of how to address it or even to understand its potential influence on vaccine acceptance. Again, individual HCPs need to navigate complex conversations with their patients and possibly their communities. Below are some suggestions that may prove helpful in formulating these conversations.

Tips for Discussion of COVID-19 Vaccines • ADDRESS CONCERNS ABOUT “NEWNESS” HCPs need to understand and be able to explain the newer vaccine platforms (mRNA, DNA, and viral vector vaccines) and to provide examples of other, older vaccines that have been developed by similar techniques. This information makes COVID-19 vaccines more familiar.

ADDRESS CONCERNs ABOUT VACCINE SAFETY HCPs need to understand and explain how vaccines are evaluated before being approved for use and how vaccine safety is monitored after vaccines are used in the population. It is important to be honest and state that potential rare and long-term effects are not yet known, but then to speak to what is from the animal and clinical trial data and to comment on background rates for rare events. Placing potential vaccine risks in the context of known COVID-19 disease risks is helpful for some patients.

Depending on the context, explain why specific high-risk groups may have been prioritized to receive the vaccine. Patients who have been prioritized may still need a strong recommendation from an HCP to accept the vaccine. An HCP recommendation is as important here as it is for acceptance of routine vaccines. As with other vaccines, many patients’ decision to accept a COVID-19 vaccine rests upon whether their HCP recommends it.

¹The Tuskegee Syphilis Study is the most infamous example of medical experimentation in Black communities in the United States. (See Brandt [1978] for details.) Numerous examples of medical experimentation on Indigenous peoples are available. For example, a 12-year trial of an experimental bacille Calmette-Guérin vaccine for tuberculosis was conducted on Cree and Nakoda Oyadabi infants in Saskatchewan during the 1930s. (See Lux [2016] for details.)

Address implicit or overt racism and systemic discrimination in the medical system and create culturally safe health care spaces. HCPs need to be aware of the legacy of discrimination, racism, and medical experimentation and the distrust it fosters in some communities. While SARS-CoV-2 has critically highlighted fractures in our health care system for minority and marginalized communities, addressing these underlying issues goes beyond addressing vaccine hesitancy and is clearly needed for all types of medical care in these communities.

EMPHASIZE THE IMPORTANCE OF KEEPING UP TO DATE WITH OTHER ROUTINE VACCINES DURING THE COVID-19 PANDEMIC These vaccines include but are not limited to seasonal influenza vaccine and the childhood primary vaccination series.

FURTHER READING

Vaccine Hesitancy

- AMERICAN ACADEMY OF PEDIATRICS: Vaccine hesitant parents. Available at www.aap.org/en-us/advocacy-and-policy/aap-health-initiatives/immunizations/Pages/vaccine-hesitant-parents.aspx. Accessed October 23, 2020.
- DESTEFANO F et al: Principal controversies in vaccine safety in the United States. *Clin Infect Dis* 69:726, 2019.
- DUDLEY MZ et al: The state of vaccine safety science: Systematic reviews of the evidence. *Lancet Infect Dis* 20:e80, 2020.
- IMMUNIZATION ACTION COALITION: For healthcare professionals. Available at www.immunize.org. Accessed October 23, 2020.
- IMMUNIZATION ACTION COALITION: For the public: Vaccine information you need. Available at vaccineinformation.org. Accessed October 23, 2020.
- JAMISON AM et al: Vaccine-related advertising in the Facebook Ad Archive. *Vaccine* 38:512, 2020.
- LEASK J et al: Communicating with parents about vaccination: A framework for health professionals. *BMC Pediatr* 12:154, 2012.
- MACDONALD N et al: Vaccine hesitancy: Definition, scope and determinants. *Vaccine* 33:4161, 2015.
- WORLD HEALTH ORGANIZATION: Vaccine hesitancy survey questions related to SAGE vaccine hesitancy. Available at www.who.int/immunization/programmes_systems/Survey_Questions_Hesitancy.pdf. Accessed October 23, 2020.
- WORLD HEALTH ORGANIZATION: Improving vaccination demand and addressing hesitancy. Available at www.who.int/immunization/programmes_systems/vaccine_hesitancy/en/. Accessed October 23, 2020.
- WORLD HEALTH ORGANIZATION: Missed opportunities for vaccination (MOV) strategy. Available at www.who.int/immunization/programmes_systems/policies_strategies/MOV/en/. Accessed October 23, 2020.

COVID-19 Vaccine Hesitancy

- BRANDT AM: Racism and research: The case of the Tuskegee Syphilis Study. *Hastings Cent Rep* 8:21, 1978.
- CENTERS FOR DISEASE CONTROL AND PREVENTION: Vaccinate with confidence: Strategy to reinforce confidence in Covid-19 vaccines. Available at www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html. Accessed April 5, 2021.
- LURIE N et al: Developing Covid-19 vaccines at pandemic speed. *N Engl J Med* 382:21, 2020.
- LUX MK: Separate beds: A history of Indian hospitals in Canada, 1920s–1980s. Toronto, University of Toronto Press, 2016.
- MOSBY I et al: Medical experimentation and the roots of COVID-19 vaccine hesitancy among Indigenous Peoples in Canada. *CMAJ* 193:E381, 2021.

4

Decision-Making in Clinical Medicine

Daniel B. Mark, John B. Wong



Practicing medicine at its core requires making decisions. What makes medical practice so difficult is not only the specialized technical knowledge required but also the intrinsic uncertainty that surrounds each decision. Mastering the technical aspects of medicine alone, unfortunately, does not ensure a mastery of the practice of medicine. Sir William Osler's familiar quote "Medicine is a science of uncertainty and an art of probability" captures well this complex duality. Although the science of medicine is often taught as if the mechanisms of the human body operate with Newtonian predictability, every aspect of medical practice is infused with an element of irreducible uncertainty that the clinician ignores at her peril. Although deeply rooted in science, more than 100 years after the practice of medicine took its modern form, it remains at its core a craft, to which individual doctors bring varying levels of skill and understanding. With the exponential growth in medical literature and other technical information and an ever-increasing number of testing and treatment options, twenty-first century physicians who seek excellence in their craft must master a more diverse and complex set of skills than any of the generations that preceded them. This chapter provides an introduction to three of the pillars upon which the craft of modern medicine rests: (1) expertise in clinical reasoning (what it is and how it can be developed); (2) rational diagnostic test use and interpretation; and (3) integration of the best available research evidence with clinical judgment in the care of individual patients (evidence-based medicine [EBM]).

BRIEF INTRODUCTION TO CLINICAL REASONING

Clinical Expertise Defining "clinical expertise" remains surprisingly difficult. Chess has an objective ranking system based on skill and performance criteria. Athletics, similarly, have ranking systems to distinguish novices from Olympians. But in medicine, after physicians complete training and pass the boards (or get recertified), no tests or benchmarks are used to identify those who have attained the highest levels of clinical performance. At each institution, there are often a few "elite" clinicians who are known for their "special problem-solving prowess" when particularly difficult or obscure cases have baffled everyone else. Yet despite their skill, even such master clinicians typically cannot explain their exact processes and methods, thereby limiting the acquisition and dissemination of the expertise used to achieve their impressive results. Furthermore, clinical virtuosity appears not to be generalizable, e.g., an expert on hypertrophic cardiomyopathy may be no better (and possibly worse) than a first-year medical resident at diagnosing and managing a patient with neutropenia, fever, and hypotension.

Broadly construed, clinical expertise encompasses not only cognitive dimensions involving the integration of disease knowledge with verbal and visual cues and test interpretation but also potentially the complex fine-motor skills necessary for invasive procedures and tests. In addition, "the complete package" of expertise in medicine requires effective communication and care coordination with patients and members of the medical team. Research on medical expertise remains sparse overall and mostly centered on diagnostic reasoning, so in this chapter, we focus primarily on the cognitive elements of clinical reasoning.

Because clinical reasoning occurs in the heads of clinicians, objective study of the process is difficult. One research method used for this area asks clinicians to "think out loud" as they receive increments of clinical information in a manner meant to simulate a clinical encounter. Another research approach focuses on how doctors should reason diagnostically, to identify remediable "errors," rather than on how they actually do reason. Much of what is known about clinical

reasoning comes from empirical studies of nonmedical problem-solving behavior. Because of the diverse perspectives contributing to this area, with important contributions from cognitive psychology, medical education, behavioral economics, sociology, informatics, and decision sciences, no single integrated model of clinical reasoning exists, and not infrequently, different terms and reasoning models describe similar phenomena.

Intuitive Versus Analytic Reasoning A useful contemporary model of reasoning, the dual-process theory distinguishes two general conceptual modes of thinking as fast or slow. *Intuition* (System 1) provides rapid effortless judgments from memorized associations using pattern recognition and other simplifying "rules of thumb" (i.e., heuristics). For example, a very simple pattern that could be useful in certain situations is "black woman plus hilar adenopathy equals sarcoid." Because no effort is involved in recalling the pattern, the clinician is often unable to say how those judgments were formulated. In contrast, *Analysis* (System 2), the other form of reasoning in the dual-process model, is slow, methodical, deliberative, and effortful. A student might read about causes of hilar adenopathy and from that list (e.g., [Chap. 66](#)), identify diseases more common in black women or examine the patient for skin or eye findings that occur with sarcoid. These dual processes, of course, represent two exemplars taken from the cognitive continuum. They provide helpful descriptive insights but very little guidance in how to develop expertise in clinical reasoning. How these idealized systems interact in different decision problems, how experts use them differently from novices, and when their use can lead to errors in judgment remain the subject of study and considerable debate.

Pattern recognition, an important part of System 1 reasoning, is a complex cognitive process that appears largely effortless. One can recognize people's faces, the breed of a dog, an automobile model, or a piece of music from just a few notes within milliseconds without necessarily being able to articulate the specific features that prompted the recognition. Analogously, experienced clinicians often recognize familiar diagnostic patterns very quickly. The key here is having a large library of stored patterns that can be rapidly accessed. In the absence of an extensive stored repertoire of diagnostic patterns, students (as well as experienced clinicians operating outside their area of expertise and familiarity) often must use the more laborious System 2 analytic approach along with more intensive and comprehensive data collection to reach the diagnosis.

The following brief patient scenarios illustrate three distinct patterns associated with hemoptysis that experienced clinicians recognize without effort:

- A 46-year-old man presents to his internist with a chief complaint of hemoptysis. An otherwise healthy, nonsmoker, he is recovering from an apparent viral bronchitis. This presentation pattern suggests that the small amount of blood-streaked sputum is due to acute bronchitis, so that a chest x-ray provides sufficient reassurance that a more serious disorder is absent.
- In the second scenario, a 46-year-old patient who has the same chief complaint but with a 100-pack-year smoking history, a productive morning cough with blood-streaked sputum, and weight loss fits the pattern of carcinoma of the lung. Consequently, along with the chest x-ray, the clinician obtains a sputum cytology examination and refers this patient for a chest CT scan.
- In the third scenario, the clinician hears a soft diastolic rumbling murmur at the apex on cardiac auscultation in a 46-year-old patient with hemoptysis who immigrated from a developing country and orders an echocardiogram as well, because of possible pulmonary hypertension from suspected rheumatic mitral stenosis.

Pattern recognition by itself is not, however, sufficient for secure diagnosis. Without deliberative systematic reflection, undisciplined pattern recognition can result in premature closure: mistakenly jumping to the conclusion that one has the correct diagnosis before all the relevant data are in. A critical second step, therefore, even when the diagnosis seems obvious, is *diagnostic verification*: considering whether

the diagnosis adequately accounts for the presenting symptoms and signs and can explain all the ancillary findings. The following case based on a real clinical encounter provides an example of premature closure. A 45-year-old man presents with a 3-week history of a “flulike” upper respiratory infection (URI) including dyspnea and a productive cough. The emergency department (ED) clinician pulled out a “URI assessment form,” which defines and standardizes the information gathered. After quickly acquiring the requisite structured examination components and noting in particular the absence of fever and a clear chest examination, the physician prescribed a cough suppressant for acute bronchitis and reassured the patient that his illness was not serious. Following a sleepless night at home with significant dyspnea, the patient developed nausea and vomiting and collapsed. He was brought back to the ED in cardiac arrest and was unable to be resuscitated. His autopsy showed a posterior wall myocardial infarction (MI) and a fresh thrombus in an atherosclerotic right coronary artery. What went wrong? Presumably, the ED clinician felt that the patient was basically healthy (one can be misled by the way the patient appears on examination—a patient that does not appear “sick” may be incorrectly assumed to have an innocuous illness). So, in this case, the physician, upon hearing the overview of the patient from the triage nurse, elected to use the URI assessment protocol even before starting the history, closing consideration of the broader range of possibilities and associated tests required to confirm or refute these possibilities. In particular, by concentrating on the abbreviated and focused URI protocol, the clinician failed to elicit the full dyspnea history, which was precipitated by exertion and accompanied by chest heaviness and relieved by rest, suggesting a far more serious disorder.

Heuristics or rules of thumb are a part of the intuitive system. These cognitive shortcuts provide a quick and easy path to reaching conclusions and making choices, but when used improperly, they can lead to errors. Two major research programs have studied heuristics in a mostly nonmedical context and have reached very different conclusions about the value of these cognitive tools. The “heuristics and biases” program focuses on how these mental shortcuts can lead to incorrect judgments. So far, however, little evidence exists that educating physicians and other decision makers to watch for the >100 cognitive biases identified to date has had any effect on the rate of diagnostic errors. In contrast, the “fast and frugal heuristics” research program explores how and when relying on simple heuristics can produce good decisions. Although many heuristics have relevance to clinical reasoning, only four will be mentioned here.

When diagnosing patients, clinicians usually develop diagnostic hypotheses based on the similarity of that patient’s symptoms, signs, and other data to their mental representations (memorized patterns) of the disease possibilities. In other words, clinicians pattern match to identify the diagnoses that share the most similar findings to the patient at hand. This cognitive shortcut is called the representativeness heuristic. Consider a patient with hypertension who has headache, palpitations, and diaphoresis. Based on the representativeness heuristic, clinicians might judge pheochromocytoma to be quite likely given this classic presenting symptom triad suggesting pheochromocytoma. Doing so, however, would be incorrect given that other causes of hypertension are much more common than pheochromocytoma and this triad of symptoms can occur in patients who do not have it. Thus, clinicians using the representativeness heuristic may overestimate the likelihood of a particular disease based on the presence of representative symptoms and signs, failing to account for its low underlying prevalence (i.e., the prior, or pretest, probabilities). Conversely, atypical presentations of common diseases may lead to underestimating the likelihood of a particular disease. Thus, inexperience with a specific disease and with the breadth of its presentations may also lead to diagnostic delays or errors, e.g., diseases that affect multiple organ systems, such as sarcoid or tuberculosis, may be particularly challenging to diagnose because of the many different patterns they may manifest.

A second commonly used cognitive shortcut, the availability heuristic, involves judgments based on how easily prior similar cases or outcomes can be brought to mind. For example, a clinician may recall

a case from a morbidity and mortality conference in which an elderly patient presented with painless dyspnea of acute onset and was evaluated for a pulmonary cause but was eventually found to have acute MI, with the diagnostic delay likely contributing to the development of ischemic cardiomyopathy. If the case was associated with a malpractice accusation, such examples may be even more memorable. Errors with the availability heuristic arise from several sources of recall bias. Rare catastrophic outcomes become memorable cases with a clarity and force disproportionate to their likelihood for future diagnosis—for example, a patient with a sore throat eventually found to have leukemia or a young athlete with leg pain subsequently found to have an osteosarcoma—and those publicized in the media or recently experienced are, of course, easier to recall and therefore more influential on clinical judgments.

The third commonly used cognitive shortcut, the anchoring heuristic (also called conservatism or stickiness), involves insufficiently adjusting the initial probability of disease up (or down) following a positive (or negative test) when compared with Bayes’ theorem, i.e., sticking to the initial diagnosis. For example, a clinician may still judge the probability of coronary artery disease (CAD) to be high despite a negative exercise perfusion test and go on to cardiac catheterization (see “Measures of Disease Probability and Bayes’ Rule,” below).

The fourth heuristic states that clinicians should use the simplest explanation possible that will adequately account for the patient’s symptoms and findings (Occam’s razor or, alternatively, the simplicity heuristic). Although this is an attractive and often used principle, it is important to remember that no biologic basis for it exists. Errors from the simplicity heuristic include premature closure leading to the neglect of unexplained significant symptoms or findings.

For complex or unfamiliar diagnostic problems, clinicians typically resort to analytic reasoning processes (System 2) and proceed methodically using the *hypothetico-deductive model of reasoning*. Based on the patient’s stated reasons for seeking medical attention, clinicians develop an initial list of diagnostic possibilities in *hypothesis generation*. During the history of the present illness, the initial hypotheses evolve in *diagnostic refinement* as emerging information is tested against the mental models of the diseases being considered with diagnoses increasing and decreasing in likelihood or even being dropped from consideration as the working hypotheses of the moment. These mental models often generate additional questions that distinguish the diagnostic possibilities from one another. The focused physical examination contributes to further distinguishing the working hypotheses. Is the spleen enlarged? How big is the liver? Is it tender? Are there any palpable masses or nodules? *Diagnostic verification* involves testing the adequacy (whether the diagnosis accounts for all symptoms and signs) and coherency (whether the signs and symptoms are consistent with the underlying pathophysiologic causal mechanism) of the working diagnosis. For example, if the enlarged and quite tender liver felt on physical examination is due to acute hepatitis (the hypothesis), then certain specific liver function tests will be markedly elevated (the prediction). Should the tests come back normal, the hypothesis may have to be discarded and others reconsidered.

Although often neglected, negative findings are as important as positive ones because they reduce the likelihood of the diagnostic hypotheses under consideration. Chest discomfort that is not provoked or worsened by exertion and not relieved by rest in an active patient lowers the likelihood that chronic ischemic heart disease is the underlying cause. The absence of a resting tachycardia and thyroid gland enlargement reduces the likelihood of hyperthyroidism in a patient with paroxysmal atrial fibrillation.

The acuity of a patient’s illness may override considerations of prevalence and the other issues described above. “*Diagnostic imperatives*” recognize the significance of relatively rare but potentially catastrophic conditions if undiagnosed and untreated. For example, clinicians should consider aortic dissection routinely as a possible cause of acute severe chest discomfort. Although the typical presenting symptoms of dissection differ from those of MI, dissection may mimic MI, and because it is far less prevalent and potentially fatal if mistreated, diagnosing dissection remains a challenging diagnostic imperative (Chap. 280).

Clinicians taking care of acute, severe chest pain patients should explicitly and routinely inquire about symptoms suggestive of dissection, measure blood pressures in both arms for discrepancies, and examine for pulse deficits. When these are all negative, clinicians may feel sufficiently reassured to discard the aortic dissection hypothesis. If, however, the chest x-ray shows a possible widened mediastinum, the hypothesis should be reinstated and an appropriate imaging test ordered (e.g., thoracic computed tomography [CT] scan or transesophageal echocardiogram). In nonacute situations, the prevalence of potential alternative diagnoses should play a much more prominent role in diagnostic hypothesis generation.

Cognitive scientists studying the thought processes of expert clinicians have observed that clinicians group data into packets, or “chunks,” that are stored in short-term or “working memory” and manipulated to generate diagnostic hypotheses. Because short-term memory is limited (classically humans can accurately repeat a list of 7 ± 2 numbers read to them), the number of diagnoses that can be actively considered in hypothesis-generating activities is similarly limited. For this reason, the cognitive shortcuts discussed above play a key role in the generation of diagnostic hypotheses, many of which are discarded as rapidly as they are formed, thereby demonstrating that the distinction between analytic and intuitive reasoning is an arbitrary and simplistic, but nonetheless useful, representation of cognition.

Research into the hypothetico-deductive model of reasoning has had difficulty identifying the elements of the reasoning process that distinguish experts from novices. This has led to a shift from examining the problem-solving process of experts to analyzing the organization of their knowledge for pattern matching as exemplars, prototypes, and illness scripts. For example, diagnosis may be based on the resemblance of a new case to patients seen previously (exemplars). As abstract mental models of disease, prototypes incorporate the likelihood of various disease features. Illness scripts include risk factors, pathophysiology, and symptoms and signs. Experts have a much larger store of exemplar and prototype cases, an example of which is the visual long-term memory of experienced radiologists. However, clinicians do not simply rely on literal recall of specific cases but have constructed elaborate conceptual networks of memorized information or models of disease to aid in arriving at their conclusions (illness scripts). That is, expertise involves an enhanced ability to connect symptoms, signs, and risk factors to one another in meaningful ways; relate those findings to possible diagnoses; and identify the additional information necessary to confirm the diagnosis.

No single theory accounts for all the key features of expertise in medical diagnosis. Experts have more knowledge about presenting symptoms of diseases and a larger repertoire of cognitive tools to employ in problem solving than nonexperts. One definition of expertise highlights the ability to make powerful distinctions. In this sense, expertise involves a working knowledge of the diagnostic possibilities and those features that distinguish one disease from another. Memorization alone is insufficient, e.g., photographic memory of a medical textbook would not make one an expert. But having access to detailed case-specific relevant information is critically important. In the past, clinicians primarily acquired clinical knowledge through their patient experiences, but now clinicians have access to a plethora of information sources. Clinicians of the future will be able to leverage the experiences of large numbers of other clinicians using electronic tools, but, as with the memorized textbook, the data alone will be insufficient for becoming an expert. Nonetheless, availability of these data removes one barrier for acquiring experience with connecting symptoms, signs, and risk factors to the possible diagnoses and identifying the additional distinguishing information necessary to confirm the diagnosis, thereby potentially facilitating the development of the working knowledge necessary for becoming an expert.

Despite all of the research seeking to understand expertise in medicine and other disciplines, it remains uncertain whether any didactic program can actually accelerate the progression from novice to expert or from experienced clinician to master clinician. Deliberate effortful practice (over an extended period of time, sometimes said to be 10 years or 10,000 practice hours) and personal coaching are two strategies often used outside medicine (e.g., music, athletics, chess) to cultivate

expertise. Their use in developing medical expertise and maintaining or enhancing it has not yet been adequately explored. Some studies in medicine suggest that the most beneficial approach to education exposes students to both the signs and symptoms of specific diseases (disease pattern recognition) and, in addition, the lists of diseases that can present with specific symptoms and signs (differential diagnosis). Active learning opportunities useful for those in training include developing a personal learning system, e.g., systematically reflecting on diagnostic processes used (metacognition) and following-up to identify diagnoses and treatments for patients in their care.

■ DIAGNOSTIC VERSUS THERAPEUTIC DECISION-MAKING

The modern ideal of medical therapeutic decision-making is to “personalize” treatment recommendations. In the abstract, personalizing treatment involves combining the best available evidence about what works with an individual patient’s unique features (e.g., risk factors, genomics, and comorbidities) and his or her preferences and health goals to craft an optimal treatment recommendation with the patient. Operationally, two different and complementary levels of personalization are possible: individualizing the risk of harm and benefit for the options being considered based on the specific patient characteristics (precision medicine), and personalizing the therapeutic decision process by incorporating the patient’s preferences and values for the possible health outcomes. This latter process is sometimes referred to as shared decision-making and typically involves clinicians sharing their knowledge about the options and the associated consequences and trade-offs and patients sharing their health goals (e.g., avoiding a short-term risk of dying from coronary artery bypass grafting to see their grandchild get married in a few months).

Individualizing the evidence about therapy **does not** mean relying on physician impressions of benefit and harm from their personal experience. Because of small sample sizes and rare events, the chance of drawing erroneous causal inferences from one’s own clinical experience is very high. For most chronic diseases, therapeutic effectiveness is only demonstrable statistically in large patient populations. It would be incorrect to infer with any certainty, for example, that treating a hypertensive patient with angiotensin-converting enzyme (ACE) inhibitors necessarily prevented a stroke from occurring during treatment, or that an untreated patient would definitely have avoided their stroke had they been treated. For many chronic diseases, a majority of patients will remain event free regardless of treatment choices; some will have events regardless of which treatment is selected; and those who avoided having an event through treatment cannot be individually identified. Blood pressure lowering, a readily observable surrogate endpoint, does not have a tightly coupled relationship with strokes prevented. Consequently, in most situations, demonstrating therapeutic effectiveness cannot rely simply on observing the outcome of an individual patient but should instead be based on large groups of patients carefully studied and properly analyzed.

Therapeutic decision-making, therefore, should be based on the best available evidence from clinical trials and well-done outcome studies. Trustworthy clinical practice guidelines that synthesize such evidence offer normative guidance for many testing and treatment decisions. However, all guidelines recognize that “one size fits all” recommendations may not apply to individual patients. Increased research into the heterogeneity of treatment effects seeks to understand how best to adjust group-level clinical evidence of treatment harms and benefits to account for the absolute level of risks faced by subgroups and even by individual patients, using, for example, validated clinical risk scores.

■ NONCLINICAL INFLUENCES ON CLINICAL DECISION-MAKING

More than three decades of research on variations in clinician practice patterns has identified important nonclinical forces that shape clinical decisions. These factors can be grouped conceptually into three overlapping categories: (1) factors related to an individual physician’s practice, (2) factors related to practice setting, and (3) factors related to payment systems.

Factors Related to Practice Style To ensure that necessary care is provided at a high level of quality, physicians fulfill a key role in medical care by serving as the patient's advocate. Factors that influence performance in this role include the physician's knowledge, training, and experience. Clearly, physicians cannot practice EBM if they are unfamiliar with the evidence. As would be expected, specialists generally know the evidence in their field better than do generalists. Beyond published evidence and practice guidelines, a major set of influences on physician practice can be subsumed under the general concept of "practice style." The practice style serves to define norms of clinical behavior. Differing practice styles may be based on training, personal experience, and medical evidence. Beliefs about effectiveness of different therapies and preferred patterns of diagnostic test use are examples of different facets of a practice style. For example, cardiologists evaluating patients with lower risk chest pain symptoms often conceptualize their primary diagnostic objective as maximizing the detection of ischemia. For this reason, they may strongly favor stress imaging. Internists caring for the same patients may be more comfortable with initial use of exercise ECG testing without imaging. This latter practice style focuses less on ischemia detection and more on following guideline recommendations that indicate no outcome advantage for stress imaging in this context. Cardiologist may also favor a more liberal use of coronary angiography and revascularization in patients with stable ischemic symptoms relative to general internists.

Beyond the patient's welfare, physician perceptions about the risk of a malpractice suit resulting from either an erroneous decision or a bad outcome may drive clinical decisions and create a practice referred to as defensive medicine. This practice involves ordering tests and therapies with very small marginal benefits, ostensibly to preclude future criticism should an adverse outcome occur. With conscious or unconscious awareness of a connection to the risk of litigation or to payment, however, over time, such patterns of care may become accepted as part of the practice norm, thereby perpetuating their overuse, e.g., annual cardiac exercise testing in asymptomatic patients.

Practice Setting Factors Factors in this category relate to work systems including tasks and workflow (interruptions, inefficiencies, workload), technology (poor design or implementation, errors in use, failure, misuse), organizational characteristics (e.g., culture, leadership, staffing, scheduling), and the physical environment (e.g., noise, lighting, layout). *Physician-induced demand* is a term that refers to the repeated observation that once medical facilities and technologies become available to physicians, they will find ways to use them. Other environmental factors that can influence decision-making include the local availability of specialists for consultations and procedures; "high-tech" advanced imaging or procedure facilities such as MRI machines and proton beam therapy centers; and fragmentation of care.

Payment Systems Economic incentives are closely related to the other two categories of practice-modifying factors. Financial issues can exert both stimulatory and inhibitory influences on clinical practice. Historically, physicians are paid on a fee-for-service, capitation, or salary basis. In fee-for-service, physicians who do more get paid more, thereby encouraging overuse, consciously or unconsciously. When fees are reduced (discounted reimbursement), clinicians tend to increase the number of services provided to maintain revenue. Capitation, in contrast, provides a fixed payment per patient per year to encourage physicians to consider a global population budget in managing individual patients and ideally reducing the use of interventions with small marginal benefit. To discourage volume-based excessive utilization, fixed salary compensation plans pay physicians the same regardless of the clinical effort expended but may provide an (unintended) incentive to see fewer patients. In recognition of the nonsustainability of continued growth in medical expenditures and the opportunity costs associated with that (funds that might be more beneficially applied to education, energy, social welfare, or defense), current efforts seek to transition to a value-based payment system to reduce overuse and to reflect benefit. Work to define how to actually tie payment to value has mostly focused so far on "pay for performance" models. High-quality

clinical trial evidence for the effectiveness of these models is still mostly lacking.

■ INTERPRETATION OF DIAGNOSTIC TESTS

Despite impressive technological advances in medicine over the past century, uncertainty still abounds and challenges all aspects of medical decision-making. Compounding this challenge, massive information overload characterizes modern medicine. Clinicians on average subscribe to seven journals, presenting them with >2500 new articles each year, and need access to 2 million pieces of information to practice medicine. Of course, to be useful, this information must be sifted for quality and examined for applicability for integration into patient-specific care. Although computers appear to offer an obvious solution both for information management and for quantification of medical care uncertainties, many practical problems remain to be solved before computerized decision support can be routinely incorporated into the clinical reasoning process in a way that demonstrably improves the quality of care. For the present, understanding the nature of diagnostic test information can help clinicians become more efficient users of such data. The next section reviews select concepts related to diagnostic testing.

■ DIAGNOSTIC TESTING: MEASURES OF TEST ACCURACY

The purpose of performing a test on a patient is to reduce uncertainty about the patient's diagnosis or prognosis in order to facilitate appropriate management. Although diagnostic tests commonly refer to laboratory (e.g., blood count) or imaging tests or procedures (e.g., colonoscopy or bronchoscopy), any information that changes a provider's understanding of the patient's problem qualifies as a diagnostic test. Thus, even the history and physical examination can be considered as diagnostic tests. In clinical medicine, it is common to reduce the results of a test to a dichotomous outcome, such as positive or negative, normal or abnormal. Although this simplification often suppresses useful information (such as the degree of abnormality), it facilitates illustrating some important principles of test interpretation that are described below.

The accuracy of any diagnostic test is assessed relative to a "gold standard," where a positive gold standard test defines the patients who have disease and a negative test securely rules out disease (**Table 4-1**). Characterizing the diagnostic performance of a new test requires identifying an appropriate population (ideally, patients representative of those in whom the new test would be used) and applying both the new and the gold standard tests to all subjects. Biased estimates of test performance occur when diagnostic accuracy is defined using an inappropriate population or one in which gold standard determination of disease status is incomplete. The accuracy of the new test in distinguishing disease from health is determined relative to the gold standard results and summarized in four estimates. The sensitivity or true-positive rate reflects how well the new test identifies patients with disease. It is the proportion of patients with disease (defined by the gold standard) who have a positive test. The proportion of patients with disease who have a negative test is the false-negative rate, calculated as $1 - \text{sensitivity}$. The specificity, or true-negative rate, reflects how well

TABLE 4-1 Measures of Diagnostic Test Accuracy

TEST RESULT	DISEASE STATUS	
	PRESENT	ABSENT
Positive	True positives (TP)	False positives (FP)
Negative	False negatives (FN)	True negatives (TN)
Test Characteristics in Patients with Disease		
True-positive rate (sensitivity) = TP/(TP + FN)		
False-negative rate = FN/(TP + FN) = 1 – true-positive rate		
Test Characteristics in Patients without Disease		
True-negative rate (specificity) = TN/(TN + FP)		
False-positive rate = FP/(TN + FP) = 1 – true-negative rate		

the new test correctly identifies patients without disease. It is the proportion of patients without disease (defined by the gold standard) who have a negative test. The proportion of patients without disease who have positive test is the false-positive rate, calculated as $1 - \text{specificity}$. In theory, a perfect test would be one with a sensitivity of 100% and a specificity of 100% and would completely distinguish patients with disease from those without it. A useful mnemonic to help remember the somewhat paradoxical relationship between what the test is best at technically versus what it is most useful for clinically is: a test with a very high sensitivity (Sn) when *negative* (N) helps *rule out* (out) disease ($SnNout$), and a test with a very high specificity (Sp) when *positive* (P) helps *rule in* (in) disease ($SpPin$).

Calculating sensitivity and specificity requires selection of a threshold value or cut point above which the test is considered “positive.” Making the cut point “stricter” (e.g., raising it) lowers sensitivity but improves specificity, while making it “laxer” (e.g., lowering it) raises sensitivity but lowers specificity. This dynamic trade-off between more accurate identification of subjects with disease versus those without disease is often displayed graphically as a receiver operating characteristic (ROC) curve (Fig. 4-1) by plotting sensitivity (y axis) versus $1 - \text{specificity}$ (x axis). Each point on the curve represents a potential cut point with an associated sensitivity and specificity value. The area under the ROC curve often is used as a quantitative measure of the information content of a test. Values range from 0.5 (no diagnostic information from testing at all; the test is equivalent to flipping a coin) to 1.0 (perfect test). The choice of cut point should in theory reflect the relative harms and benefits of treatment for those without versus those with disease. For example, if treatment was safe with substantial benefit, then choosing a high-sensitivity cut point (upper right of the ROC curve) for a low-risk test may be appropriate (e.g., phenylketonuria in newborns), but if treatment had substantial risk for harm, then choosing a high-specificity cut point (lower left of the ROC curve) may be appropriate (e.g., chemotherapy for cancer). The choice of cut point may also depend on the prevalence of disease, with low prevalence placing a greater emphasis on the harms of false-positive tests (e.g., HIV testing in marriage applicants) or the harms of false-negative tests (e.g., HIV testing in blood donors).

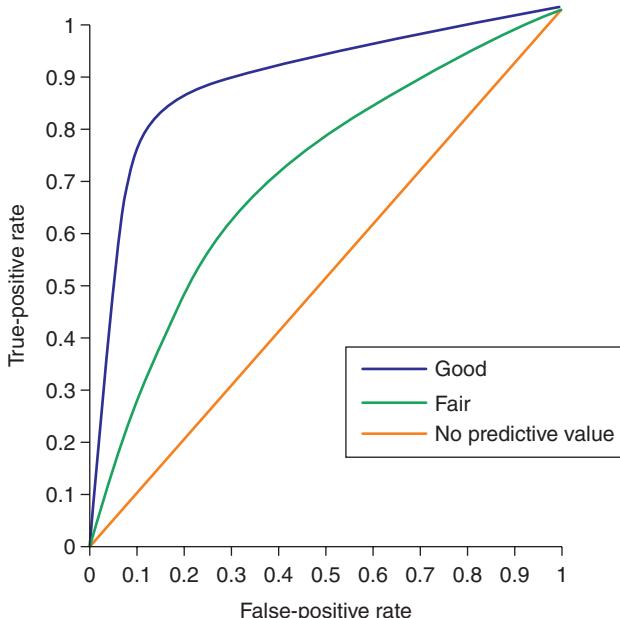


FIGURE 4-1 Each receiver operating characteristic (ROC curve) illustrates a trade-off that occurs between improved test sensitivity (accurate detection of patients with disease) and improved test specificity (accurate detection of patients without disease), as the test value defining when the test turns from “negative” to “positive” is varied. A 45° line would indicate a test with no predictive value (sensitivity = specificity at every test value). The area under each ROC curve is a measure of the information content of the test. Thus, a larger ROC area signifies increased diagnostic accuracy.

MEASURES OF DISEASE PROBABILITY AND BAYES’ RULE

In the absence of perfect tests, the true disease state of the patient remains uncertain after every test. Bayes’ rule provides a way to quantify the revised uncertainty using simple probability mathematics (and thereby avoid anchoring bias). It calculates the *posttest probability*, or likelihood of disease after a test result, from three parameters: the pretest probability of disease, the test sensitivity, and the test specificity. The *pretest probability* is a quantitative estimate of the likelihood of the diagnosis before the test is performed and is usually estimated from the prevalence of the disease in the underlying population (if known) or clinical context (e.g., age, sex, and type of chest pain). For some common conditions, such as CAD, existing nomograms and statistical models generate estimates of pretest probability that account for history, physical examination, and test findings. The posttest probability (also called the predictive value of the test, see below) is a recalibrated statement of the probability of the diagnosis, accounting for both pretest probability and test results. For the probability of disease following a positive test (i.e., positive predictive value), Bayes’ rule is calculated as:

$$\text{Posttest probability} = \frac{\text{Pretest probability} \times \text{test sensitivity}}{\text{Pretest probability} \times \text{test sensitivity} + (1 - \text{Pretest probability}) \times (\text{false-positive test rate})}$$

For example, consider a 64-year-old woman with atypical chest pain who has a pretest probability of 0.50 and a “positive” diagnostic test result (assuming test sensitivity = 0.90 and specificity = 0.90).

$$\text{Posttest probability} = \frac{(0.50)(0.90)}{(0.50)(0.90) + (0.50)(0.10)} = 0.90$$

The term *predictive value* has often been used as a synonym for the posttest probability. Unfortunately, clinicians commonly misinterpret reported predictive values as intrinsic measures of test accuracy rather than calculated probabilities. Studies of diagnostic test performance compound the confusion by calculating predictive values from the same sample used to measure sensitivity and specificity. Such calculations are misleading unless the test is applied subsequently to populations with exactly the same disease prevalence. For these reasons, the term *predictive value* is best avoided in favor of the more descriptive posttest probability following a positive or a negative test result.

The nomogram version of Bayes’ rule (Fig. 4-2) helps us to understand at a conceptual level how it estimates the posttest probability of disease. In this nomogram, the impact of the diagnostic test result is summarized by the likelihood ratio, which is defined as the ratio of the probability of a given test result (e.g., “positive” or “negative”) in a patient with disease to the probability of that result in a patient without disease, thereby providing a measure of how well the test distinguishes those with from those without disease.

The *likelihood ratio for a positive test* is calculated as the ratio of the true-positive rate to the false-positive rate (or sensitivity/[$1 - \text{specificity}$]). For example, a test with a sensitivity of 0.90 and a specificity of 0.90 has a likelihood ratio of $0.90/(1 - 0.90)$, or 9. Thus, for this hypothetical test, a “positive” result is 9 times more likely in a patient with the disease than in a patient without it. Most tests in medicine have likelihood ratios for a positive result between 1.5 and 20. Higher values are associated with tests that more substantially increase the posttest likelihood of disease. A very high likelihood ratio positive (>10) usually implies high specificity, so a positive high specificity test helps “rule in” disease (the “SpPin” mnemonic introduced earlier). If sensitivity is excellent but specificity is less so, the likelihood ratio positive will be reduced substantially (e.g., with a 90% sensitivity but a 55% specificity, the likelihood ratio positive is 2.0).

The corresponding *likelihood ratio for a negative test* is the ratio of the false-negative rate to the true-negative rate (or [$1 - \text{sensitivity}$]/specificity).

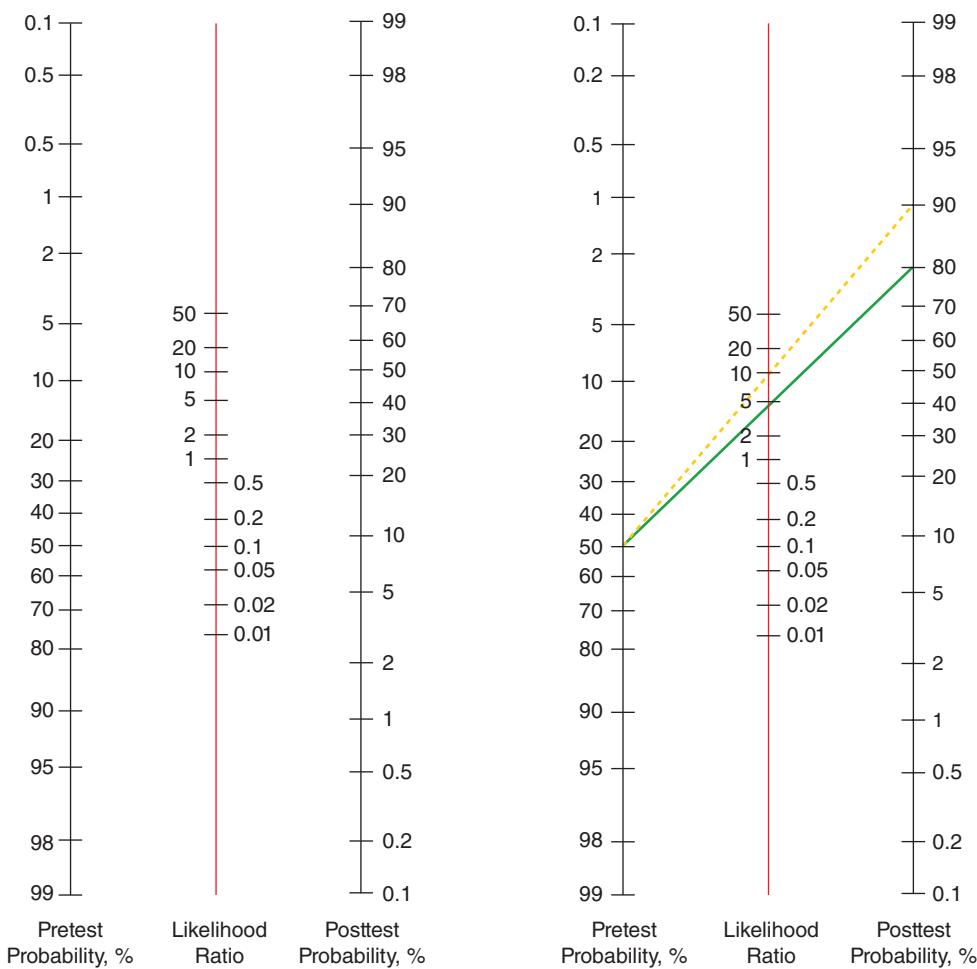


FIGURE 4-2 Nomogram version of Bayes' theorem used to predict the posttest probability of disease (right-hand scale) using the pretest probability of disease (left-hand scale) and the likelihood ratio for a positive or a negative test (middle scale). See text for information on calculation of likelihood ratios. To use, place a straightedge connecting the pretest probability and the likelihood ratio and read off the posttest probability. The right-hand part of the figure illustrates the value of a positive exercise treadmill test (likelihood ratio 4, green line) and a positive exercise thallium single-photon emission CT perfusion study (likelihood ratio 9, broken yellow line) in a patient with a pretest probability of coronary artery disease of 50%. (*Adapted from Centre for Evidence-Based Medicine: Likelihood ratios. Available at <http://www.cebm.net/likelihood-ratios/>.*)

Lower likelihood ratio negative values more substantially lower the posttest likelihood of disease. A very low likelihood ratio negative (falling below 0.10) usually implies high sensitivity, so a negative high sensitivity test helps “rule out” disease (the SnNout mnemonic). The hypothetical test considered above with a sensitivity of 0.9 and a specificity of 0.9 would have a likelihood ratio for a negative test result of $(1 - 0.9)/0.9$, or 0.11, meaning that a negative result is about one-tenth as likely in patients with disease than in those without disease (or about 10 times more likely in those without disease than in those with disease).

■ APPLICATIONS TO DIAGNOSTIC TESTING IN CAD

Consider two tests commonly used in the diagnosis of CAD: an exercise treadmill and an exercise single-photon emission CT (SPECT) myocardial perfusion imaging test (Chap. 241). A positive treadmill ST-segment response has an average sensitivity of ~60% and an average specificity of ~75%, yielding a likelihood ratio positive of 2.4 ($0.60/[1 - 0.75]$) (consistent with modest discriminatory ability because it falls between 2 and 5). For a 41-year-old man with nonanginal pain and a 10% pretest probability of CAD, the posttest probability of disease after a positive result rises to only ~30%. For a 60-year-old woman with typical angina and a pretest probability of CAD of 80%, a positive test result raises the posttest probability of disease to ~95%.

In contrast, exercise SPECT myocardial perfusion test is more accurate for diagnosis of CAD. For simplicity, assume that the finding of a reversible exercise-induced perfusion defect has both a sensitivity and

a specificity of 90% (a bit higher than reported), yielding a likelihood ratio for a positive test of 9.0 ($0.90/[1 - 0.90]$) (consistent with intermediate discriminatory ability because it falls between 5 and 10). For the same 10% pretest probability patient, a positive test raises the probability of CAD to 50% (Fig. 4-2). However, despite the differences in posttest probabilities between these two tests (30 vs 50%), the more accurate test may not improve diagnostic likelihood enough to change patient management (e.g., decision to refer to cardiac catheterization) because the more accurate test has only moved the physician from being fairly certain that the patient did not have CAD to a 50:50 chance of disease. In a patient with a pretest probability of 80%, exercise SPECT test raises the posttest probability to 97% (compared with 95% for the exercise treadmill). Again, the more accurate test does not provide enough improvement in posttest confidence to alter management, and neither test has improved much on what was known from clinical data alone.

In general, positive results with an accurate test (e.g., likelihood ratio for a positive test of 10) when the pretest probability is low (e.g., 20%) do not move the posttest probability to a range high enough to rule in disease (e.g., 80%). In screening situations, pretest probabilities are often particularly low because patients are asymptomatic. In such cases, specificity becomes especially important. For example, in screening first-time female blood donors without risk factors for HIV, a positive test raised the likelihood of HIV to only 67% despite a specificity of 99.995%

because the prevalence was 0.01%. Conversely, with a high pretest probability, a negative test may not rule out disease adequately if it is not sufficiently sensitive. Thus, the largest change in diagnostic likelihood following a test result occurs when the clinician is most uncertain (i.e., pretest probability between 30 and 70%). For example, in patients with a pretest probability for CAD of 50%, a positive exercise treadmill test moves the posttest probability to 80% and a positive exercise SPECT test moves it to 90% (Fig. 4-2).

As presented above, Bayes' rule employs a number of important simplifications that should be considered. First, few tests provide only “positive” or “negative” results. Many tests have multidimensional outcomes (e.g., extent of ST-segment depression, exercise duration, and exercise-induced symptoms with exercise testing). Although Bayes' theorem can be adapted to this more detailed test result format, it is computationally more complex to do so. Similarly, when multiple sequential tests are performed, the posttest probability may be used as the pretest probability to interpret the second test. However, this simplification assumes conditional independence—that is, that the results of the first test do not affect the likelihood of the second test result—and this is often not true.

Finally, many texts assert that sensitivity and specificity are prevalence-independent parameters of test accuracy. This statistically useful assumption, however, is often incorrect. A treadmill exercise test, for example, has a sensitivity of ~30% in a population of patients with one-vessel CAD, whereas its sensitivity in patients with severe three-vessel CAD approaches 80%. Thus, the best estimate of sensitivity

to use in a particular decision may vary, depending on the severity of disease in the local population. A hospitalized, symptomatic, or referral population typically has a higher prevalence of disease and, in particular, a higher prevalence of more advanced disease than does an outpatient population. Consequently, test sensitivity will likely be higher in hospitalized patients and test specificity higher in outpatients.

■ STATISTICAL PREDICTION MODELS

Bayes' rule, when used as presented above, is useful in studying diagnostic testing concepts, but predictions based on multivariable statistical models can more accurately address these more complex problems by simultaneously accounting for additional relevant patient characteristics. In particular, these models explicitly account for multiple, even possibly overlapping, pieces of patient-specific information and assign a relative weight to each on the basis of its unique independent contribution to the prediction in question. For example, a logistic regression model to predict the probability of CAD ideally considers all the relevant independent factors from the clinical examination and diagnostic testing and their relative importance instead of the limited data that clinicians can manage in their heads or with Bayes' rule. However, despite this strength, prediction models are usually too complex computationally to use without a calculator or computer. Guideline-driven treatment recommendations based on statistical prediction models available online, e.g., the American College of Cardiology/American Heart Association risk calculator for primary prevention with statins and the CHA₂DS₂-VASC calculator for anticoagulation for atrial fibrillation, have generated more widespread usage. When electronic health records (EHRs) will provide sufficient platform support to allow for routine use of predictive models in clinical practice and increase their impact on clinical encounters and outcomes remains uncertain.

One reason for limited clinical use is that, to date, only a handful of prediction models have been validated sufficiently (for example, Wells criteria for pulmonary embolism; **Table 4-2**). The importance of independent validation in a population separate from the one used to develop the model cannot be overstated. An unvalidated prediction model should be viewed with the skepticism appropriate for any new drug or medical device that has not had rigorous clinical trial testing.

When statistical survival models in cancer and heart disease have been compared directly with clinicians' predictions, the survival models have been found to be more consistent, as would be expected, but not always more accurate. On the other hand, comparison of clinicians with websites and apps that generate lists of possible diagnoses to help patients with self-diagnosis found that physicians outperformed the currently available programs. For students and less-experienced clinicians, the biggest value of diagnostic decision support may be in extending diagnostic possibilities and triggering "rational override," but their impact on knowledge, information-seeking, and problem-solving needs additional research.

TABLE 4-2 Wells Clinical Prediction Rule for Pulmonary Embolism (PE)

CLINICAL FEATURE	POINTS
Clinical signs of deep-vein thrombosis	3
Alternative diagnosis is less likely than PE	3
Heart rate >100 beats/min	1.5
Immobilization ≥3 days or surgery in previous 4 weeks	1.5
History of deep-vein thrombosis or pulmonary embolism	1.5
Hemoptysis	1
Malignancy (with treatment within 6 months) or palliative	1
INTERPRETATION	
Score >6.0	High
Score 2.0–6.0	Intermediate
Score <2.0	Low

FORMAL DECISION SUPPORT TOOLS

■ DECISION SUPPORT SYSTEMS

Over the past 50 years, many attempts have been made to develop computer systems to aid clinical decision-making and patient management. Conceptually, computers offer several levels of potentially useful support for clinicians. At the most basic level, they provide ready access to vast reservoirs of information, which may, however, be quite difficult to sort through to find what is needed. At higher levels, computers can support care management decisions by making accurate predictions of outcome, or can simulate the whole decision process, and provide algorithmic guidance. Computer-based predictions using Bayesian or statistical regression models inform a clinical decision but do not actually reach a "conclusion" or "recommendation." Machine learning methods are being applied to pattern recognition tasks such as the examination of skin lesions and the interpretation of x-rays. Artificial intelligence (AI) systems attempt to simulate or replace human reasoning with a computer-based analogue. Natural language processing allows the system to access and process large amounts of data, both from the EHR and from the medical literature. To date, such approaches have achieved only limited success. The most prominent example, IBM's Watson program, introduced publicly in 2011, has yet to produce persuasive evidence of clinical decision support utility. Reminder or protocol-directed systems do not make predictions but use existing algorithms, such as guidelines or appropriate utilization criteria, to direct clinical practice. In general, however, decision support systems have so far had little impact on practice. Reminder systems built into EHRs have shown the most promise, particularly in correcting drug dosing and promoting adherence to guidelines. Checklists may also help avoid or reduce errors.

■ DECISION ANALYSIS

Compared with the decision support methods discussed earlier, decision analysis represents a normative prescriptive approach to decision-making in the face of uncertainty. Its principal application is in complex decisions. For example, public health policy decisions often involve *trade-offs* in length versus quality of life, benefits versus resource use, population versus individual health, and *uncertainty* regarding efficacy, effectiveness, and adverse events as well as *values* or preferences regarding mortality and morbidity outcomes.

One recent analysis using this approach involved the optimal screening strategy for breast cancer, which has remained controversial, in part because a randomized controlled trial to determine when to begin screening and how often to repeat screening mammography is impractical. In 2016, the National Cancer Institute-sponsored Cancer Intervention and Surveillance Network (CISNET) examined eight strategies differing by whether to initiate mammography screening at age 40, 45, or 50 years and whether to screen annually, biennially, or annually for women in their forties and biennially thereafter (hybrid). The six simulation models found biennial strategies to be the most efficient for average-risk women. Biennial screening for 1000 women from age 50–74 years versus no screening avoided seven breast cancer deaths. Screening annually from age 40–74 years avoided three additional deaths but required 20,000 additional mammograms and yielded 1988 more false-positive results. Factors that influenced the results included patients with a 2–4-fold higher risk for developing breast cancer in whom annual screening from age 40–74 years yielded similar benefits as biennial screening from age 50–74. For average-risk patients with moderate or severe comorbidities, screening could be stopped earlier, at age 66–68 years.

This analysis involved six models that reproduced epidemiologic trends and a screening trial result, accounted for digital technology and treatments advances, and considered quality of life, risk factors, breast density, and comorbidity. It provided novel insights into a public health problem in the absence of a randomized clinical trial and helped weigh the pros and cons of such a health policy recommendation. Although such models have been developed for selected clinical problems, their benefit and application to individual real-time clinical management has yet to be demonstrated.

DIAGNOSIS AS AN ELEMENT OF QUALITY OF CARE

High-quality medical care begins with accurate diagnosis. The incidence of diagnostic errors has been estimated by a variety of methods including postmortem examinations, medical record reviews, and medical malpractice claims, with each yielding complementary but different estimates of this quality of care patient-safety problem. In the past, diagnostic errors tended to be viewed as a failure of individual clinicians. The modern view is that they are mostly a system of care deficiencies. Current estimates suggest that nearly everyone will experience at least one diagnostic error in their lifetime, leading to mortality, morbidity, unnecessary tests and procedures, costs, and anxiety.

Solutions to the “diagnostic errors as a system of care” problem have focused on system-level approaches, such as decision support and other tools integrated into EHRs. The use of checklists has been proposed as a means of reducing some of the cognitive errors discussed earlier in the chapter, such as premature closure. While checklists have been shown to be useful in certain medical contexts, such as operating rooms and intensive care units, their value in preventing diagnostic errors that lead to patient adverse events remains to be shown.

EVIDENCE-BASED MEDICINE

Clinical medicine is defined traditionally as a practice combining medical knowledge (including scientific evidence), intuition, and judgment in the care of patients (**Chap. 1**). Evidence-based medicine (EBM) updates this construct by placing much greater emphasis on the processes by which clinicians gain knowledge of the most up-to-date and relevant clinical research to determine for themselves whether medical interventions alter the disease course and improve the length or quality of life. The phrase “evidence-based medicine” is now used so often and in so many different contexts that many practitioners are unaware of its original meaning. The intention of the EBM program, as described in the early 1990s by its founding proponents at McMaster University, becomes clearer through an examination of its four key steps:

1. Formulating the management question to be answered
2. Searching the literature and online databases for applicable research data
3. Appraising the evidence gathered with regard to its validity and relevance
4. Integrating this appraisal with knowledge about the unique aspects of the patient (including the patient’s preferences about the possible outcomes)

The process of searching the world’s research literature and appraising the quality and relevance of studies can be time-consuming and requires skills and training that most clinicians do not possess. In a busy clinical practice, the work required is also logistically not feasible. This has led to a focus on finding recent systematic overviews of the problem in question as a useful shortcut in the EBM process. Systematic reviews are regarded by some as the highest level of evidence in the EBM hierarchy because they are intended to comprehensively summarize the available evidence on a particular topic. To avoid the potential biases found in narrative review articles, predefined reproducible explicit search strategies and inclusion and exclusion criteria seek to find all of the relevant scientific research and grade its quality. The prototype for this kind of resource is the Cochrane Database of Systematic Reviews. When appropriate, a meta-analysis is used to quantitatively summarize the systematic review findings (discussed further below).

Unfortunately, systematic reviews are not uniformly the acme of the EBM process they were initially envisioned to be. In select circumstances, they can provide a much clearer picture of the state of the evidence than is available from any individual clinical report, but their value is less clear when only a few trials are available, when trials and observational studies are mixed, or when the evidence base is only observational. They cannot compensate for deficiencies in the underlying research available, and many are created without the requisite clinical insights. The medical literature is now flooded with systematic

reviews of varying quality and clinical utility. The peer review system has, unfortunately, not proved to be an effective arbiter of quality of these papers. Therefore, systematic reviews should be used with circumspection in conjunction with selective reading of some of the best empirical studies.

SOURCES OF EVIDENCE: CLINICAL TRIALS AND REGISTRIES

The notion of learning from observation of patients is as old as medicine itself. Over the past 50 years, physicians’ understanding of how best to turn raw observation into useful evidence has evolved considerably. Medicine has received a hard refresher lesson in this process from COVID-19 pandemic. Starting in the spring of 2020, case reports, personal and institutional anecdotal experience, and small single-center case series started appearing in the peer-reviewed literature and within months turned into a flood of confusing and often contradictory evidence. Observational reports of treatments for COVID-19 fueled the confusion. Despite >40,000 publications appearing in the first 7 months of the pandemic, an enormous amount of uncertainty around prevention, diagnosis, treatment, and prognosis of the disease remained. Many of the early 2020 publications were either small observational series or reviews of published series, neither of which can resolve the key uncertainties clinicians need to address in caring for these patients. These small observational studies often have substantial limitations in validity and generalizability, and although they may generate important hypotheses or be the first reports of adverse events or therapeutic benefit, they have no role in formulating modern standards of practice. The major tools used to develop reliable evidence consist of randomized clinical trials supplemented strategically by large (high-quality) observational registries. A registry or database typically is focused on a disease or syndrome (e.g., different types of cancer, acute or chronic CAD, pacemaker capture, or chronic heart failure), a clinical procedure (e.g., bone marrow transplantation, coronary revascularization), or an administrative process (e.g., claims data used for billing and reimbursement).

By definition, in observational data, the investigator does not control patient care. Carefully collected prospective observational data, however, can at times achieve a level of evidence quality approaching that of major clinical trial data. At the other end of the spectrum, data collected retrospectively (e.g., chart review) are limited in form and content to what previous observers recorded and may not include the specific research data being sought (e.g., claims data). Advantages of observational data include the inclusion of a broader population as encountered in practice than is typically represented in clinical trials because of their restrictive inclusion and exclusion criteria. In addition, observational data provide primary evidence for research questions when a randomized trial cannot be performed. For example, it would be difficult to randomize patients to test diagnostic or therapeutic strategies that are unproven but widely accepted in practice, and it would be unethical to randomize based on sex, racial/ethnic group, socioeconomic status, or country of residence or to randomize patients to a potentially harmful intervention, such as smoking or deliberately overeating to develop obesity.

A well-done prospective observational study of a particular management strategy differs from a well-done randomized clinical trial most importantly by its lack of protection from treatment selection bias. The use of observational data to compare diagnostic or therapeutic strategies assumes that sufficient uncertainty and heterogeneity exists in clinical practice to ensure that similar patients will be managed differently by diverse physicians. In short, the analysis assumes that a sufficient element of randomness (in the sense of disorder rather than in the formal statistical sense) exists in clinical management. In such cases, statistical models attempt to adjust for important imbalances to “level the playing field” so that a fair comparison among treatment options can be made. When management is clearly not random (e.g., all eligible left main CAD patients are referred for coronary bypass surgery), the problem may be too confounded (biased) for statistical correction, and observational data may not provide reliable evidence.

In general, the use of concurrent controls is vastly preferable to that of historical controls. For example, comparison of current surgical management of left main CAD with medically treated patients with left main CAD during the 1970s (the last time these patients were routinely treated with medicine alone) would be extremely misleading because “medical therapy” has substantially improved in the interim.

Randomized controlled clinical trials include the careful prospective design features of the best observational data studies but also include the use of random allocation of treatment. This design provides the best protection against measured and unmeasured confounding due to treatment selection bias (a major aspect of internal validity). However, the randomized trial may not have good external validity (generalizability) if the process of recruitment into the trial resulted in the exclusion of many potentially eligible subjects or if the nominal eligibility for the trial describes a very heterogeneous population.

Consumers of medical evidence need to be aware that randomized trials vary widely in their quality and applicability to practice. The process of designing such a trial often involves many compromises. For example, trials designed to gain U.S. Food and Drug Administration (FDA) approval for an investigational drug or device must fulfill regulatory requirements (such as the use of a placebo control) that may result in a trial population and design that differ substantially from what practicing clinicians would find most useful.

META-ANALYSIS

The Greek prefix *meta* signifies something at a later or higher stage of development. Meta-analysis is research that combines and summarizes the available evidence quantitatively. Although it is used to examine nonrandomized studies, meta-analysis is most useful for summarizing all available randomized trials examining a particular therapy used in a specific clinical context. Ideally, unpublished trials should be identified and included to avoid publication bias (i.e., missing “negative” trials that may not be published). Furthermore, the best meta-analyses obtain and analyze individual patient-level data from all trials rather than using only the summary data from published reports. Nonetheless, not all published meta-analyses yield reliable evidence for a particular problem, so their methodology should be scrutinized carefully to ensure proper study design and analysis. The results of a well-done meta-analysis are likely to be most persuasive if they include at least several large-scale, properly performed randomized trials. Meta-analysis can especially help detect benefits when individual trials are inadequately powered (e.g., the benefits of streptokinase thrombolytic therapy in acute MI demonstrated by ISIS-2 in 1988 were evident by the early 1970s through meta-analysis). However, in cases in which the available trials are small or poorly done, meta-analysis should not be viewed as a remedy for deficiencies in primary trial data or trial design.

Meta-analyses typically focus on summary measures of relative treatment benefit, such as odds ratios or relative risks. Clinicians should also examine what absolute risk reduction (ARR) can be expected from the therapy. A metric of absolute treatment benefit that is frequently reported is the number needed to treat (NNT) to prevent one adverse outcome event (e.g., death, stroke). NNT should not be interpreted literally as a causal statement. NNT is simply 1/ARR. For example, if a hypothetical therapy reduced mortality rates over a 5-year follow-up by 33% (the relative treatment benefit) from 12% (control arm) to 8% (treatment arm), the ARR would be $12\% - 8\% = 4\%$ and the NNT would be $1/0.04$, or 25. This does not mean literally that 1 patient benefits and 24 do not. However, it can be conceptualized as an informal measure of treatment efficiency. If the hypothetical treatment was applied to a lower-risk population, say, with a 6% 5-year mortality, the 33% relative treatment benefit would reduce absolute mortality by 2% (from 6% to 4%), and the NNT for the same therapy in this lower-risk group of patients would be 50. Although not always made explicit, comparisons of NNT estimates from different studies should account for the duration of follow-up used to create each estimate. In addition, the NNT concept assumes a homogeneity in response to treatment that may not be accurate. The NNT is simply another way of summarizing the absolute treatment difference and does not provide any unique information.

CLINICAL PRACTICE GUIDELINES

Per the 1990 Institute of Medicine definition, clinical practice guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.” This definition emphasizes several crucial features of modern guideline development. First, guidelines are created by using the tools of EBM. In particular, the core of the development process is a systematic literature search followed by a review of the relevant peer-reviewed literature. Second, guidelines usually are focused on a clinical disorder (e.g., diabetes mellitus, stable angina pectoris) or a health care intervention (e.g., cancer screening). Third, the primary objective of guidelines is to improve the quality of medical care by identifying care practices that should be routinely implemented, based on high-quality evidence and high benefit-to-harm ratios for the interventions. Guidelines are intended to “assist” decision-making, not to define explicitly what decisions should be made in a particular situation, in part because guideline-level evidence alone is never sufficient for clinical decision-making (e.g., deciding whether to intubate and administer antibiotics for pneumonia in a terminally ill individual, in an individual with dementia, or in an otherwise healthy 30-year-old mother).

Guidelines are narrative documents constructed by expert panels whose composition often is determined by interested professional organizations. These panels vary in expertise and in the degree to which they represent all relevant stakeholders. The guideline documents consist of a series of specific management recommendations, a summary indication of the quantity and quality of evidence supporting each recommendation, an assessment of the benefit-to-harm ratio for the recommendation, and a narrative discussion of the recommendations. Many recommendations simply reflect the expert consensus of the guideline panel because literature-based evidence is insufficient or absent. A recent examination of this issue in cardiovascular guidelines showed that <15% of guideline recommendations were based on the highest level of clinical trial evidence, and this proportion had not improved in 10 years despite a substantial number of trials being conducted and published. The final step in guideline construction is peer review, followed by a final revision in response to the critiques provided.

Guidelines are closely tied to the process of quality improvement in medicine through their identification of evidence-based best practices. Such practices can be used as quality indicators. Examples include the proportion of acute MI patients who receive aspirin upon admission to a hospital and the proportion of heart failure patients with a depressed ejection fraction treated with an ACE inhibitor.

CONCLUSIONS

Thirty years after the introduction of the EBM movement, it is tempting to think that all the difficult decisions practitioners face have been or soon will be solved and digested into practice guidelines and computerized reminders. However, EBM provides practitioners with an ideal rather than a finished set of tools with which to manage patients. Moreover, even with such evidence, it is always worth remembering that the response to therapy of the “average” patient represented by the summary clinical trial outcomes may not be what can be expected for the specific patient sitting in front of a provider in the clinic or hospital. In addition, meta-analyses cannot generate evidence when there are no adequate randomized trials, and most of what clinicians confront in practice will never be thoroughly tested in a randomized trial. For the foreseeable future, excellent clinical reasoning skills and experience supplemented by well-designed quantitative tools and a keen appreciation for the role of individual patient preferences in their health care will continue to be of paramount importance in the practice of clinical medicine.

FURTHER READING

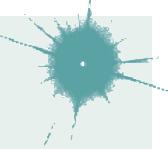
- CROSKERRY P: A universal model of diagnostic reasoning. *Acad Med* 84:1022, 2009.
- DHALIWAL G, DETSKY AS: The evolution of the master diagnostician. *JAMA* 310:579, 2013.

- FANAROFF AC et al: Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. *JAMA* 321:1069, 2019.
- HUNINK MGM et al: *Decision Making in Health and Medicine: Integrating Evidence and Values*, 2nd ed. Cambridge, Cambridge University Press, 2014.
- KAHNEMAN D: *Thinking Fast and Slow*. New York, Farrar, Straus and Giroux, 2013.
- KASSIRER JP et al: *Learning Clinical Reasoning*, 2nd ed. Baltimore, Lippincott Williams & Wilkins, 2009.
- MANDELBLATT JS et al: Collaborative modeling of the benefits and harms of associated with different U.S. breast cancer screening strategies. *Ann Intern Med* 164:215, 2016.
- MONTEIRO S et al: The 3 faces of clinical reasoning: Epistemological explorations of disparate error reduction strategies. *J Eval Clin Pract* 24:666, 2018.
- MURTHY VK et al: An inquiry into the early careers of master clinicians. *J Grad Med Educ* 10:500, 2018.
- RICHARDS JB et al: Teaching clinical reasoning and critical thinking: From cognitive theory to practical application. *Chest* 158:1617, 2020.
- ROYCE CS et al: Teaching critical thinking: A case for instruction in cognitive biases to reduce diagnostic errors and improve patient safety. *Acad Med* 94:187, 2019.
- SAPOSNIK G et al: Cognitive biases associated with medical decisions: A systematic review. *BMC Med Inform Decis Mak* 16:138, 2016.
- SCHUWIRTH LWT et al: Assessment of clinical reasoning: three evolutions of thought. *Diagnosis (Berl)* 7:191, 2020.

5

Precision Medicine and Clinical Care

The Editors



DISEASE NOSOLOGY AND PRECISION MEDICINE

Modern disease nosology arose in the late nineteenth century and represented a clear departure from the holistic, limited descriptions of disease dating to Galen. In this rubric, the definition of any disease is largely based on clinicopathologic observation. As the correlation between clinical signs and symptoms with pathoanatomy required autopsy material, diseases tended to be characterized by the end organ in which the primary syndrome was manifest and by late-stage presentations. Morgagni institutionalized this framework with the publication of *De Sedibus et Causis Morborum per Anatomam Indagatis* in 1761, in which he correlated the clinical features of patients with more than 600 autopsies at the University of Padua, demonstrating an anatomic basis for disease pathophysiology. Clinicopathologic observation served as the basis for inductive generalization coupled with the application of Occam's razor in which disease complexity was reduced to its simplest possible form. While this approach to defining human disease has held sway for over a century and facilitated the conquest of many diseases previously considered incurable, overly inclusive and simplified Oslerian diagnostics suffer from significant shortcomings. These include, but are not limited to, failure to distinguish the underlying etiology of different diseases with common pathophenotypes. For example, many different diseases can cause end-stage kidney disease or heart failure. Over time, the classification of neurodegenerative disorders or lymphomas, as well as many other diseases, is becoming more refined and precise as the underlying etiologies are identified. These distinctions are important for providing predictable prognostic information for individual patients with even highly prevalent diseases. Additionally, therapies may be ineffective owing to a lack of understanding of the often subtle molecular complexities of specific disease drivers.

Beginning in the mid-twentieth century, the era of molecular medicine offered the idealized possibility of identifying the underlying molecular basis of every disease. Using a conventional reductionist paradigm, physician-scientists explored disease mechanism at ever-increasing molecular depth, seeking the single (or limited number of) molecular cause(s) of many human diseases. Yet, as effective as this now conventional scientific approach was at uncovering many disease mechanisms, the clinical manifestations of very few diseases could be explained on the basis of a single molecular mechanism. Even knowledge of the globin β chain mutation that causes sickle cell disease does not predict the many different manifestations of the disease (stroke syndrome, painful crises, and hemolytic crisis, among others). Clearly, the profession had expected too much from oversimplified reductionism and failed to take into consideration the extraordinary biologic variety and its accompanying molecular and genetic complexity that underpin both normal and pathologic diversity. The promise of the Human Genome Project provided new tools and approaches and unleashed efforts to identify a monogenic, oligogenic, or polygenic cause for every disease (allowing for environmental modulation). Yet, once again, disappointment reigned as the pool of genomes expanded without the expected revelations (aside from rare variants). The arc of progressive reductionism (as illustrated for tuberculosis in Fig. 5-1) in refining and explaining disease reached a humbling plateau, revealing the need for new approaches to understand better the etiology, manifestations, and progression of most diseases. The stage was set for a return to holism. However, in contrast to the holism of ancient physicians, we adopted one that is integrative, taking genomic context into account in all dimensions. In the course of elaborating this complex pathobiologic landscape, disease definition must become more precise and progressively more individualized, setting the stage for what we term *precision medicine*.

Oversimplification of phenotype is a natural outgrowth of the observational scientific method. Categorizing individuals as falling into groups or clusters that are reasonably similar simplifies the task of the diagnostician and also facilitates the application of "specific" therapies more broadly. Biomedicine has been viewed as less quantitative and precise than other scientific disciplines, with biologic and pathobiologic diversity (biologic "noise") viewed as the norm. Thus, distilling such observational complexity to a fundamental group of symptoms or signs that are reasonably invariant across a group of sick individuals has served as the basis for the approach to disease and its treatment since the earliest days of medicine. This approach to diagnosis and therapy has remained in place into the twenty-first century, serving as the basis for the development of standard diagnostic tests and of broadly applied drug therapies. Targeting larger groups of patients is efficient when applied to large populations. As successful as this approach has been in advancing medical care, it is important to point out its limitations, which include significant predictive inaccuracies and sizeable segments of the disease population who do not respond to the most "effective" drugs (upward of 60% by some estimates). Clearly, a more nuanced approach to diagnosis and therapy is required to achieve better prognostic and therapeutic outcomes.

Turning first to phenotype, astute clinicians know full well the subtle and vivid differences in presentation that are often manifest among individuals with the same disease. In some cases, these differences in pathophenotype lead to new subclassifications of the disease, such as heart failure with preserved ejection fraction versus heart failure with reduced ejection fraction. Often, these relatively crude efforts at making diagnoses more precise are driven by new technologies or new ways of applying established technologies. In other cases, differences in pathophenotype are more subtle, not necessarily clinically apparent, and often driven by measures of endophenotype, such as distinctions among vasculitides facilitated by refinements in serologies or immunophenotyping. The impetus to create these subclasses of disease is largely determined by the need to improve prognosis and apply more precise and effective therapies. Based on these guiding principles, many experienced clinicians will argue—and rightly so—that they have been practicing personalized, precision medicine throughout their careers: they characterize each patient's illness in great detail, and choose therapies that respect and are guided by those individualized clinical and laboratory features, limited though they may be.

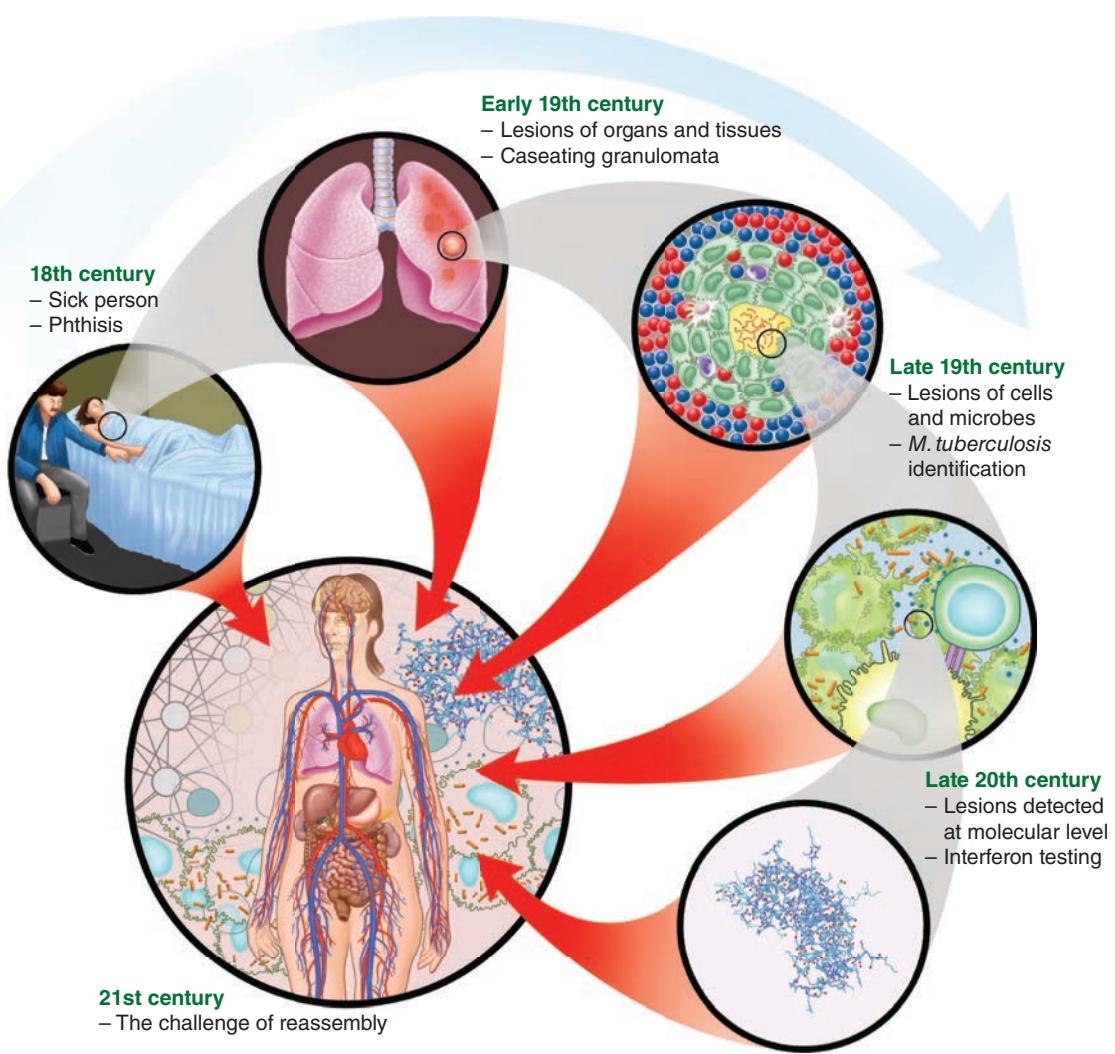


FIGURE 5-1 Arc of reductionism in medicine. (From JA Greene, J Loscalzo. Putting the patient back together—social medicine, network medicine, and the limits of reductionism. *N Engl J Med* 377:2493, 2017. Copyright © 2017 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

For many diseases, genomic variation, whether inherited or acquired, provides opportunities to refine diagnostic precision with even greater fidelity and predictive accuracy. For this reason, the field of precision medicine has now entered a new era that couples the molecular reductionism of the last century with an integrative, systems-level understanding of the basis for pathophenotype. Equally important, modern genomics has established that genomic context, sometimes referred to as modifier genes, is distinctive for each individual person; hence, understanding that context provides the insight necessary to predict how a primary disease driver or drivers may manifest a clinical pathophenotype—e.g., why some individuals with sickle cell anemia will develop stroke, while others will develop acute chest syndrome. This concept that primary genetic and/or environmental drivers of a disease differentially affect disease expression based on an individual's unique genomic context serves as the ultimate basis for much of what we denote as precision medicine.

To develop a precision medicine strategy for any disease, the clinician needs to be aware of two important, confounding principles. First, patients with different diseases can manifest similar pathophenotypes, i.e., *convergent phenotypes*. Examples of this principle include the hypertrophied myocardium found in hypertrophic cardiomyopathy, infiltrative cardiomyopathies, critical aortic stenosis, and untreated, long-standing hypertension; and the thrombotic microangiopathy found in malignant hypertension, scleroderma renal crisis, thrombotic thrombocytopenic purpura, eclampsia, and antiphospholipid syndrome. Second, patients with the same basic disease can manifest very different pathophenotypes, i.e., *divergent phenotypes* (Chap. 466). Examples of this principle include the different clinical manifestations of cystic fibrosis or sickle cell disease and the incomplete penetrance of

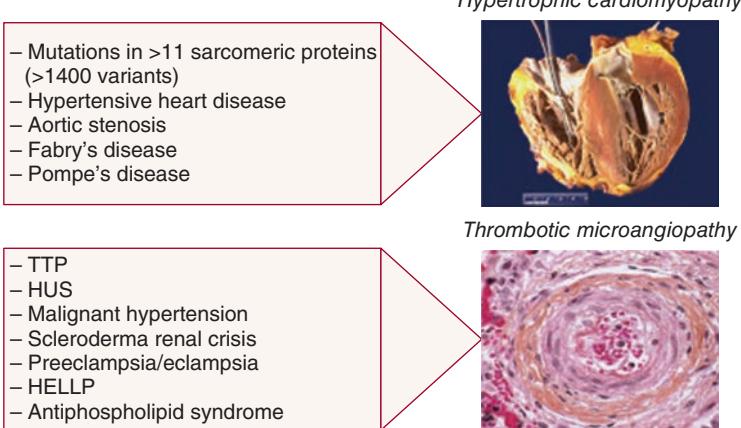
many common genetic diseases. These common presentations of different diseases and different presentations of the same disease are both a consequence of genomic context coupled with unique exposures over an individual's lifetime (Fig. 5-2). Understanding the interplay among these many complex molecular determinants of disease expression is essential for the success of precision medicine.

Given the complexity of the genomic and environmental context of an individual, one must ask the question: How precise do we need to be in order to practice effective precision medicine? Complete knowledge of a person's comprehensive genome (DNA, gene expression, mitochondrial function, proteome, metabolome, posttranslational modification of the proteome, and metagenome, among others) and quantitative assessments of environmental and social history are not possible to acquire; yet, this shortcoming does not render the general problem intractable. Owing to the fact that the molecular networks that govern phenotype are overdetermined (i.e., redundant) and that there are primary drivers of disease expression that are modified in a weighted way by other genomic features of an individual, the practice of precision medicine can be realized without complete knowledge of all dimensions of the genome. Examples of how best to realize this strategy are discussed later in this chapter.

■ REQUIREMENTS FOR PRECISION MEDICINE

The essential elements of any precision medicine effort include phenotyping, endophenotyping (defining the characteristics of a disorder that are not readily observable), and genomic profiling (Fig. 5-3). While subtle distinctions among individuals with the same disease are well known to clinicians, formalizing these nuanced differences is critical for achieving more precise phenotypes. Deep phenotyping requires a

A



B

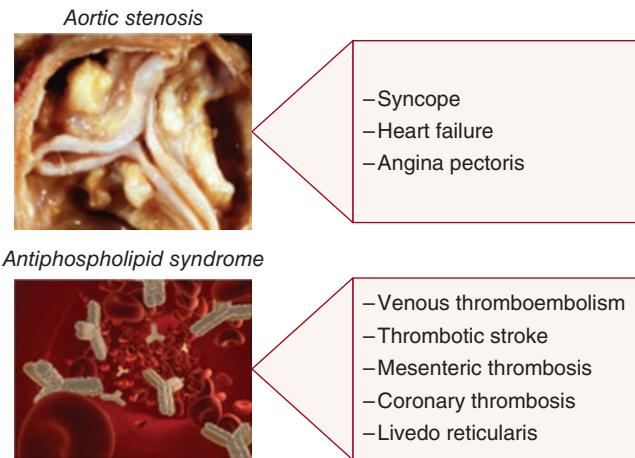


FIGURE 5-2 Convergent and divergent phenotypes. Examples of the former (**A**) include hypertrophic cardiomyopathy and thrombotic microangiopathy, and examples of the latter, and (**B**) include aortic stenosis and antiphospholipid syndrome, each of which can have several distinct clinical presentations. HELLP, hemolysis, elevated liver enzymes, and a low platelet count; HUS, hemolytic-uremic syndrome; TTP, thrombotic thrombocytopenic purpura.

detailed history, including family history and environmental exposures, as well as relevant (physiologic) functional studies and imaging, including molecular imaging where appropriate. Biochemical, immunologic, and molecular tests of body fluids provide additional detail to the overall phenotype. Importantly, these objective laboratory tests together with functional studies compose an assessment of the endophenotype (or endotype) of an individual, refining the overall discriminant power of the evaluation. One additional concept that has gained traction in recent years is the notion of orthogonal phenotyping, i.e., assessing clinical, molecular, imaging, or functional (endo)phenotypes seemingly unrelated to the clinical presentation. These features further enhance the ability to distinguish (sub) phenotypes and derive from the fact that diseases can be subtly (subclinically) manifest in organ systems different from that in which the primary symptoms or signs are expressed. While some diseases are well known to affect multiple organ systems (e.g., systemic lupus erythematosus) and in many cases involvement of those many systems is assessed at initial diagnosis, such is not the case for most other diseases. As we begin to understand the differences in the organ-specific expression of genomic variants that drive or modify disease, it is becoming increasingly apparent that orthogonal—or more appropriately, unbiased comprehensive—phenotyping should become the norm.

Genomic profiling must next be coupled to detailed phenotyping. The complex levels of genomic assessment continue to mature and include DNA sequencing (exomic, whole genome), gene expression (mRNA and protein expression), and metabolomics. In addition, the epigenome, the posttranslationally modified proteome, and the metagenome (the personal microbiome of an individual) are gaining traction as additional elements of comprehensive genomics (Chap. 483). Not all of these genomic features are yet available for clinical laboratory testing, and those that are available are largely confined to blood testing. While DNA sequencing using whole blood would generally apply to any organ-based disease,

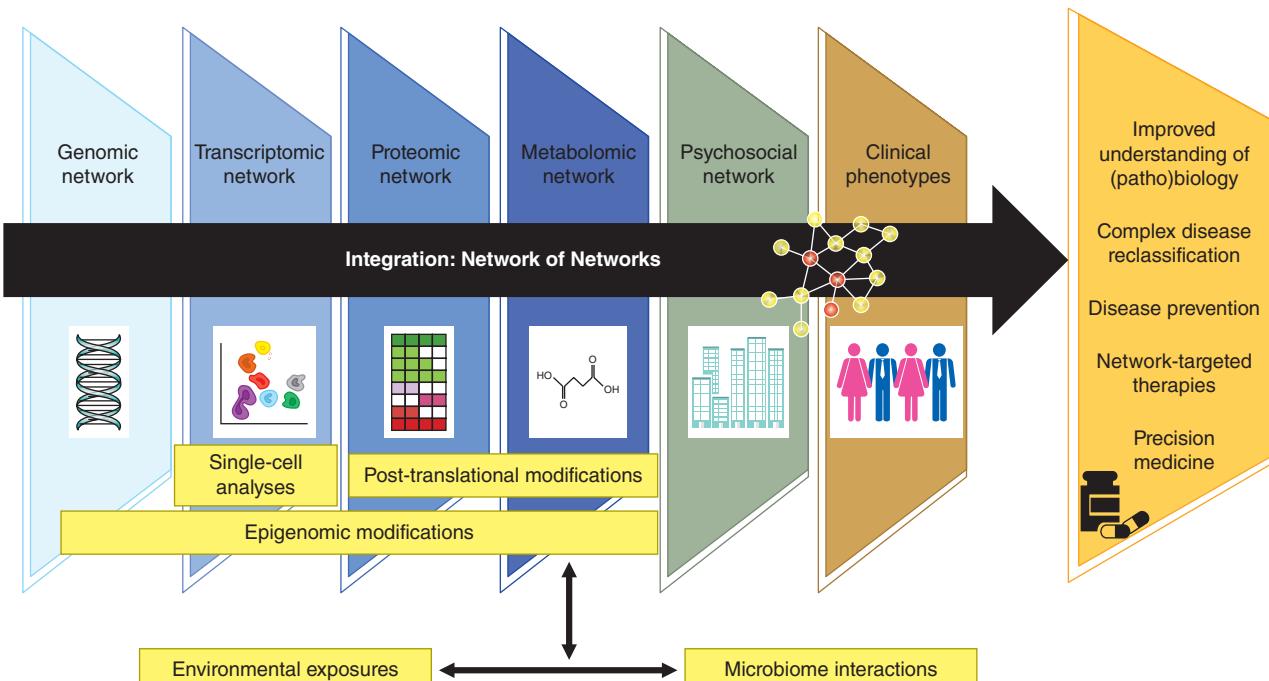


FIGURE 5-3 Universe of precision medicine. The totality of precision medicine incorporates multidimensional biologic networks, the integration of which leads to a network of networks whose components interact with each other and with environmental exposures to yield a distinctive phenotype or pathophenotype. (Reproduced with permission from LY-H Lee, J Loscalzo: Network medicine in pathobiology. *Am J Pathol* 189:1311, 2019.)

gene expression, metabolomics, and epigenetics are often tissue specific. As tissue specimens cannot always or easily be obtained from the organ of interest, attempts at correlating whole-blood mRNA, protein, or metabolite profiles with those of the involved organ are critical for precise prognostics and therapeutic choices. In many cases, systemic consequences to an organ-specific disease (e.g., systemic inflammatory responses in individuals with atherosclerosis) can be ascertained and may provide useful prognostic information or therapeutic strategies. These biomarker signatures are the subject of ongoing discovery and have provided useful guidance toward improved diagnostic precision in many diseases. However, in many diseases, the correlations between these plasma or blood markers and organ-based diseases are weak, indicating a need to analyze each condition and each resulting signature before applying it to clinical decision-making. It is important to note that one of the key determinants of the functional consequences of a genetic variant believed to drive a disease phenotype is not simply its expression in a tissue of interest but, more importantly, the coexpression of protein binding partners in that same tissue comprising specific (dys)functional pathways that govern phenotype (Fig. 5-4). An alternative strategy currently under investigation is the conversion of induced pluripotent stem cells from a patient into a cell type of interest for gene expression or metabolomics study. As rational as this approach seems from first principles, it is important to note that gene expression patterns in these induced, differentiated cell types are not completely consonant with their native counterparts, offering often limited additional information at potentially great additional expense.

While phenotype features of many chronic diseases are assessed over time, genomic features tend to be limited to single time point sampling. Time trajectories are extremely informative in precision genotyping and phenotyping, with gene expression patterns and phenotypes changing over time in different ways among different patients with the same overarching phenotype. Cost, feasible sampling frequency, predictive power, and therapeutic choices will all drive the optimal strategy for the acquisition of timed samples in any given patient; however, with continued cost reduction in genomics technologies, this limitation may be progressively mitigated and clinical application may become a reality.

One important class of diseases that does not have most of these limitations in genomic profiling is cancer. Cancers can be (and are) sampled (biopsied) frequently to monitor temporal changes in the somatically mutating oncogenome and its consequences for the limited number of well-defined oncogenic driver pathways (Chap. 68). A unique limitation of cancer in this regard, however, is that the frequency of somatic mutations over time (and, especially, with treatment) is great and the functional consequences of many of these mutations unknown. Equally important, assessment of single-cell mRNA sequencing patterns demonstrates great variability between apparently similar cells, challenging functional interpretation. Lastly, in solid tumors, stromal cells interact in a variety of ways (e.g., metabolically) with the associated malignant cells, and their gene expression signatures are also modified by the changing somatic mutational landscape of the primary malignancy. Thus, while much more information can be obtained over time in most cancer patients, the interpretation of these rich data sets continues to remain largely semi-empirical.

The possibility of identifying specific therapeutic targets remains a major goal of precision medicine. Doing so requires more than simple DNA sequencing and must include analysis of some level of gene expression, ideally in the involved organ(s). In addition to demonstrating the expression of a variant protein in the organ, one must ideally also demonstrate its functional consequences, which requires ascertaining the expression of binding partner proteins and the functional pathways they comprise. To achieve this goal, a variety of approaches have been tried, one of the most successful of which is the construction of the protein-protein interactome (the interactome), which is a comprehensive network map of the protein-protein interactions in a cell or organ of interest (Chap. 486). This template provides information on the subnetworks that govern a disease phenotype (disease modules), which can be further individualized by incorporating individual variants and differentially expressed proteins that are patient specific. This type of analysis leads to the creation of an individual “reticulome” or

reticulotype, which links the genotype to the phenotype of an individual (Fig. 5-5). Using this approach, one can identify potential drug targets in a rational way or can even repurpose existing drugs by demonstrating the proximity of a known drug target to a disease module of interest (Fig. 5-6). For example, in multicentric Castleman's disease, a disorder of unclear etiology, recognition that the PI3K/Akt/mTOR pathway is highly activated led to trials with an existing, approved drug, sirolimus. Precision medicine offers additional opportunities for optimizing the utilization of a drug by assessing the individualized pharmacogenomics of its disposition and metabolism, as demonstrated for the adverse consequences of variants in *TPMT* on azathioprine metabolism and in *CYP2C19* on clopidogrel metabolism (Chap. 68).

■ EXAMPLES OF PRECISION MEDICINE APPLICATIONS

The field of precision medicine did not appear abruptly in medical history but, rather, evolved gradually as clinicians became more aware of differences among patients with the same disease. With the advent of modern genomics, in the ideal situation, these phenotype differences can now be mapped to genotype differences. Thus, we can consider precision medicine from the perspective of the pregenomic era and the postgenomic era. Pregenomic precision medicine was applied to many diseases as therapeutic classes expanded for those disorders. A prime example of this approach is in the field of heart failure, where diuretics, digoxin, beta blockers, afterload-reducing agents, venodilators, renin-angiotensin-aldosterone inhibitors, and brain natriuretic peptide (nesiritide) are commonly used in some combination for most patients. The choice of agents is governed by the evidence basis for their use, but tailored to the primary pathophysiologic phenotypes manifest in a patient, such as congestion, hypertension, and impaired contractility. These treatments were developed in the latter half of the last century based on empiric observation, reductionist experiments of specific pathways believed to be involved in the pathophysiology, and clinical response in prospective trials. As phenotyping became more refined (e.g., echocardiographic assessments of ventricular function and tissue Doppler characterization of ventricular relaxation), the syndrome was subclassified into heart failure with reduced ejection fraction and heart failure with preserved ejection fraction, the latter of which does not respond well to any of the classes of therapeutic agents currently available. In the postgenomic era, ever more refined and detailed methods are under investigation to characterize pathophenotypes as well as genotypes, which may then be matched to the idealized combination of therapeutic classes of agents.

Pulmonary arterial hypertension is another disease for which definitive therapies straddle the pre- and postgenomic eras of precision medicine. Prior to the 1990s, there were no effective therapies for this highly morbid and lethal condition. With the advent of molecular and biochemical characterization of vascular abnormalities in individuals with established disease, however, therapies with agents that restored normal vascular function improved morbidity and mortality. These included calcium channel blockers, prostacyclin congeners, and endothelin receptor antagonists. As genomic characterization of the disease has progressed over the past two decades, there is increasing recognition of distinct genotypes that yield unique phenotypes (Chap. 283), such as the demonstration of a primarily fibrotic endophenotype governed by the (oxidized) scaffold protein NEDD-9 and its aldosterone-dependent, TGF- β -independent enhancement of collagen III expression. This approach will continue to evolve as therapies become more effective (e.g., for perivascular fibrosis) and therapeutic choices better targeted to individual patients.

Precision genomics has also led to a new classification of the dementias, conditions previously thought to have a single cause with varied clinical expression. These disorders can now be categorized based on the genes and pathways involved and the site where aggregated proteins first form and then spread in the nervous system. For example, the varied clinical presentations of frontotemporal dementia, including progressive aphasia, behavioral disturbances, and dementia with amyotrophic lateral sclerosis, can now be linked to specific genotypes and susceptible cells (Chap. 432). In prion diseases, the clinical phenotype

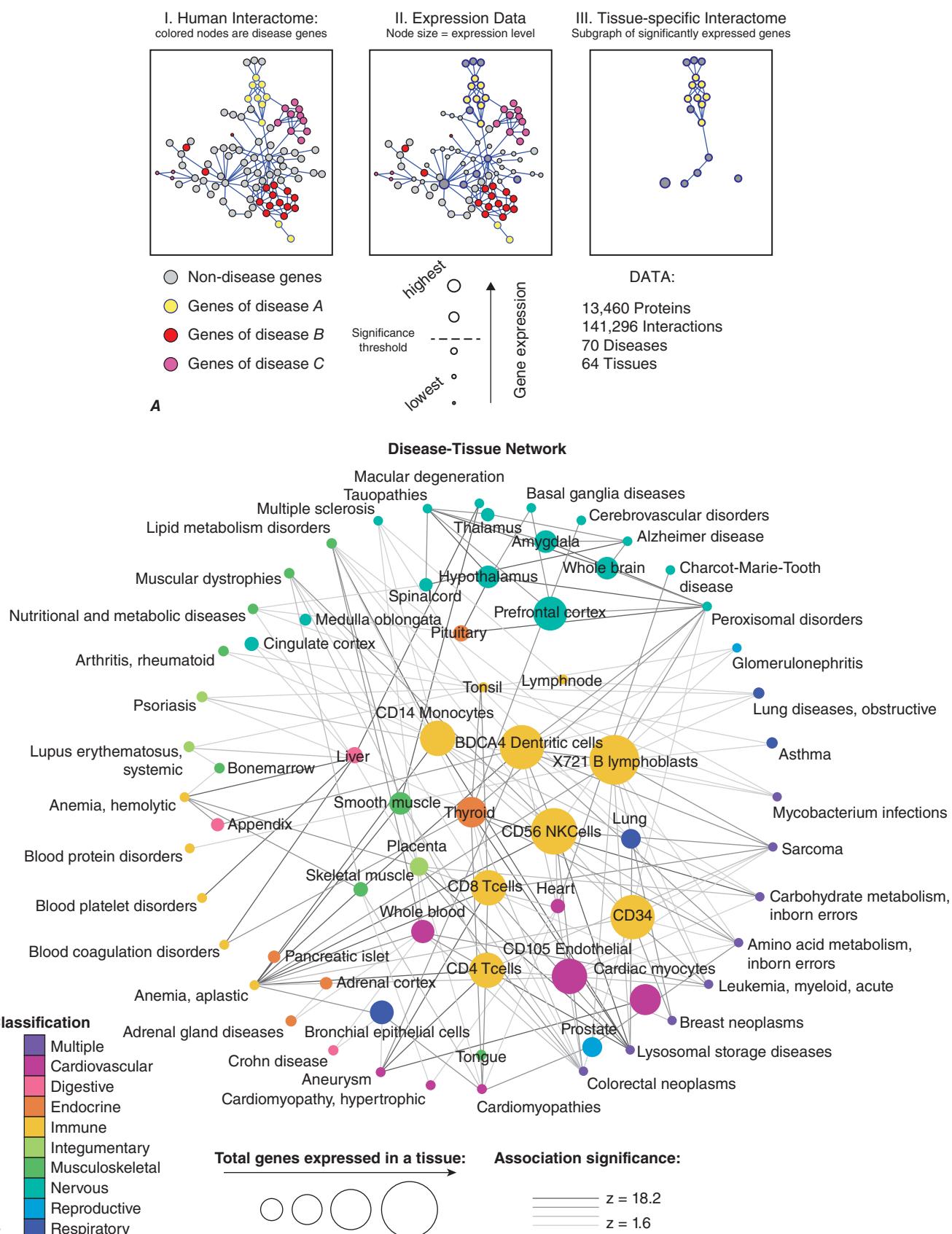


FIGURE 5-4 Gene expression and phenotype. **A.** The human protein-protein interactome is constructed, and a specific disease module is identified (I); gene expression within this module is ascertained (II); and the tissue specificity of gene expression is determined (III). This analysis leads to a reduction of the total number of disease module genes that govern phenotype in a specific organ, which is a reflection of the specific pathway (or pathways) that is (or are) expressed in their functional entirety in that tissue. **B.** A disease-tissue bipartite network is constructed wherein specific tissues are placed within the circle and linked to diseases shown on the circumference. Nodes are colored according to tissue classification, the sizes of nodes are proportional to the total number of genes expressed in them, and the widths (shades) of the lines or edges correspond to the significance of the associations with specific diseases. (From M Kitsak et al: Tissue Specificity of Human Disease Module. *Sci Rep* 6: 35241, 2016, Figure 4.)

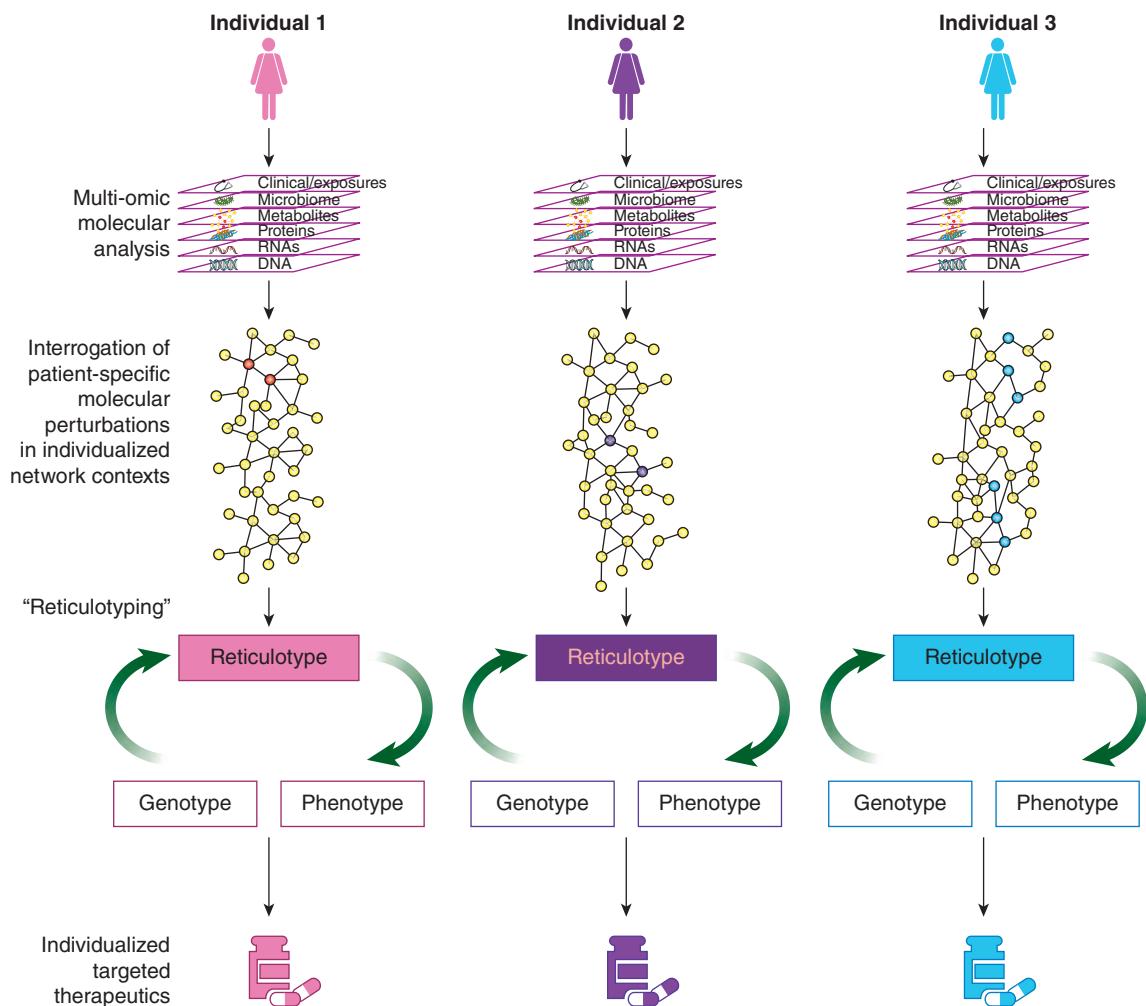


FIGURE 5-5 Reticulotyping. Patient-specific genotype-phenotype relationships by multiomic network structures are depicted for three individuals. Each individual's unique molecular perturbations (genetic variants, differentially expressed genes) are examined within the context of the subject's unique integrative biologic network or reticulome derived from these multiomic analyses. These unique reticulotypes then serve as the basis for patient-specific, precision therapies. (Reproduced with permission from LYH Lee, J Loscalzo: *Network Medicine in Pathobiology*. Am J Pathol 189:1311, 2019.)

is determined by specific germline mutations present in the prion protein ([Chap. 438](#)). Discovery of autoantibodies against aquaporin-4 (AQP-4) and myelin oligodendrocyte glycoprotein (MOG) has allowed neuromyelitis optica, previously considered a multiple sclerosis-like disorder, to be classified as a separate entity requiring different treatment ([Chap. 445](#)). Similarly, in myasthenia gravis, the identification of novel autoantibodies now permits stratification and a more finely tuned precision approach to therapy ([Chap. 448](#)).

Precision medicine approaches to cancers have, of course, become the prime example of the opportunity that this strategy offers. In the pregenomic era, chemotherapy was widely used with variable success despite continued efforts to characterize the molecular features of the specific tumors and their semi-empiric responses to specific chemotherapeutic agents. As cancer genome sequencing evolved, however, it became apparent that there are a limited number of oncogenic pathways (<20) that are represented in the great majority of malignancies, without regard for the organ in which the disease was primarily manifest. These genomic signatures served as a template for precisely targeted therapies that have led to dramatic changes in response to treatment, including, for example, imatinib (and congeners) for Bcr-Abl tyrosine kinase activity in chronic myelogenous leukemia, erlotinib for EGFR-mutant non-small cell lung cancers, and ibrutinib for Bruton tyrosine kinase in chronic lymphocytic leukemia, among many others.

As exciting as these approaches have been, there are at least three primary challenges associated with precision therapeutics that are unique to cancer: (1) the mutational landscape continues to evolve as the disease progresses, and therapy often (if not invariably) leads to selection for resistant clones; (2) the likelihood that any cancer can be definitively cured by any single agent, no matter its exquisite precision, is quite limited, necessitating

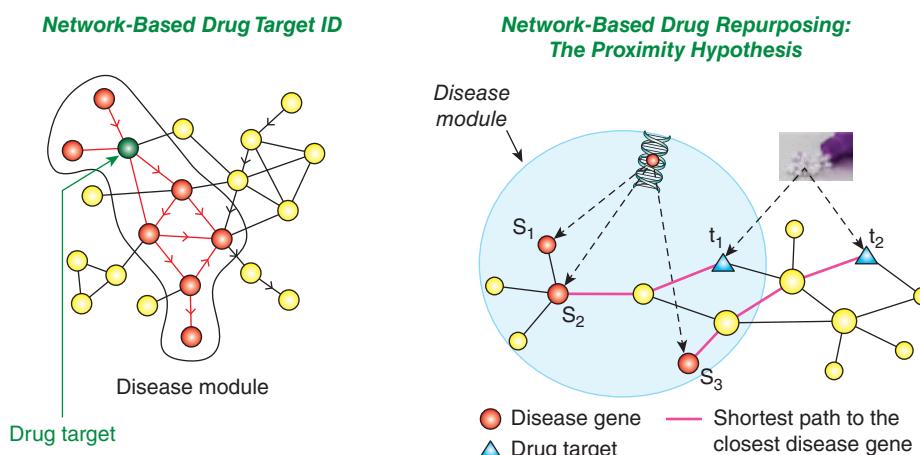


FIGURE 5-6 Network-based precision drug repurposing. (Adapted from F Cheng et al: A genome-wide positioning systems network algorithm for *in silico* drug repurposing. Nat Commun 10:3476, 2019.)

the development of rational polypharmaceutical approaches that take into account alternative pathways that achieve the same oncogenic goals as the primary targeted pathway, complicating drug development; and (3) there is marked genomic heterogeneity in many malignancies arguing that targeting a specific pathway—even with multiple drugs—may not ultimately succeed over the long term owing to the continued and heterogeneous evolution of the genomic landscape within a tumor within a patient. Despite these serious shortcomings, the application of progressively more refined and precisely targeted therapies used alone and in combination, such as with immune modulators, continues to offer great promise for the treatment of these diseases. In some ways, these approaches in cancer mirror earlier strategies in the treatment of infectious diseases in which the identification of the causative organism and its sensitivity to potential antimicrobials allows precision approaches to treatment. Combinatorial antimicrobial treatments represent an effective strategy to address acquired resistance. These diagnostic and therapeutic strategies can be applied without detailed knowledge of personalized responses to the infection or treatment (aside from serious adverse effects) with good outcomes in most cases. Yet, individuals do respond differently to specific infections and their treatments, possibly driven by different endophenotypes (e.g., different inflammatory responses), suggesting that more precise knowledge of these precise mechanistic differences may yield improved prognosis and therapeutic approaches. As with cancer, immune modulation, particularly for immune exhaustion in chronic infections, represents a new frontier, again amenable to the personalized, precise analyses described above.

THE FUTURE OF PRECISION MEDICINE

Precision medicine clearly holds great promise for the future of the practice of medicine. For precision medicine to continue to evolve successfully, however, several requirements will need to be met. First, both deeply refined personal phenotypic data and genomic data are essential as the information with which precision analysis is performed. These data sets are quite large and require sufficient storage for analysis, especially for individuals in whom time trajectories are acquired (as should be the case for every person). Equally important, the analytical methods required to extract useful information from these data sets are evolving and themselves quite complex. While great progress has been made in genomics and biochemical testing, our ability to capture meaningful immunologic endophenotypes and environmental exposures is limited by comparison. Machine learning and artificial (auxiliary) intelligence methods will be essential for extracting optimal information from these data sets, which include not only pathways that can be uniquely targeted therapeutically but also individualized genomic or phenotypic signatures that are highly predictive of outcome, with or without therapy. Gathering sufficient information on the “normal” segments of the population is also required to ensure appropriate comparison data sets for optimal prediction.

Second, phenotyping must continue to expand and become dimensionally richer. The phenotypic features included in this data gathering must incorporate not only data relevant to the clinical presentation but also orthogonal phenotypic data that may yield useful information on disease trajectory or preclinical disease markers. Personal device data, environmental exposure history, social network interactions, and health system data will all be incorporated increasingly in defining phenotype and will require great efforts on the part of the medical informatics community to harmonize data sets, standardize data collection, and optimize/standardize data analysis (Fig. 5-7).

Third, perhaps the greatest challenge to making precision medicine the standard approach to illness will be to determine the minimal data set required to predict outcome and response to therapy. Gathering data is comparatively simple; however, analyzing it to eliminate redundant information in these overdetermined biologic systems, weighting the determinants of an outcome, and using the data as phenomic/genomic signatures that are easier to collect than comprehensive, unbiased data sets are the ideal goals—a major challenge, but not insurmountable. Rapidly evolving machine learning and artificial intelligence strategies will also be essential for maximal success.

To return to the question of how precise precision medicine needs to be in order to be useful, please refer to Fig. 5-8 where the approaches

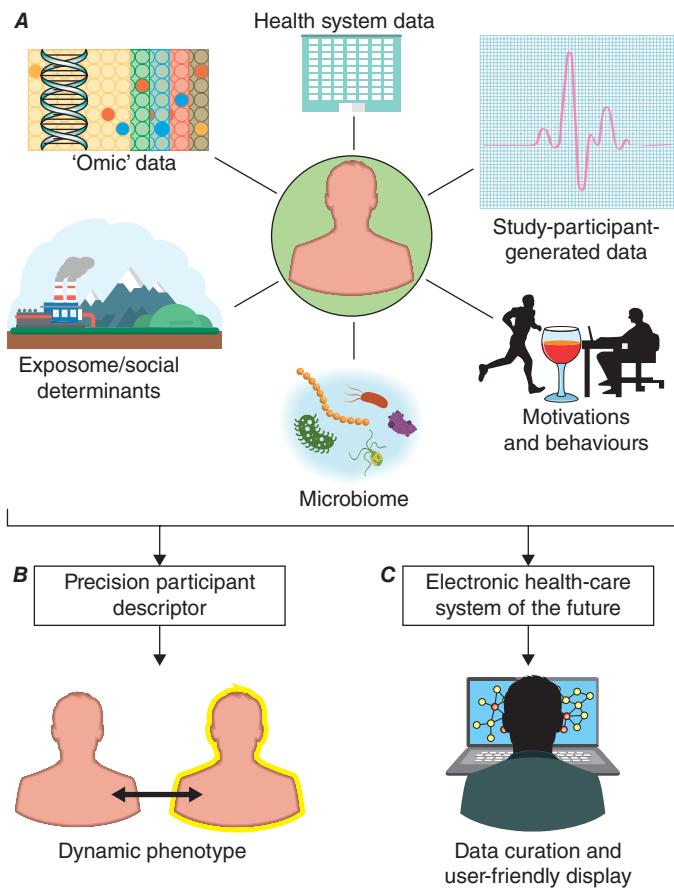


FIGURE 5-7 Big data in precision medicine. **A**, Six dimensions by which individuals may be characterized in the precision medicine era are described. **B**, The precision participant descriptor integrates the data from these six dimensions and varies over time. **C**, The electronic medical record increasingly must evolve to provide curated precision data in a user-friendly way. (Reproduced with permission from EM Antman, J Loscalzo: Precision medicine in cardiology. *Nat Rev Cardiol* 13:591, 2016.)

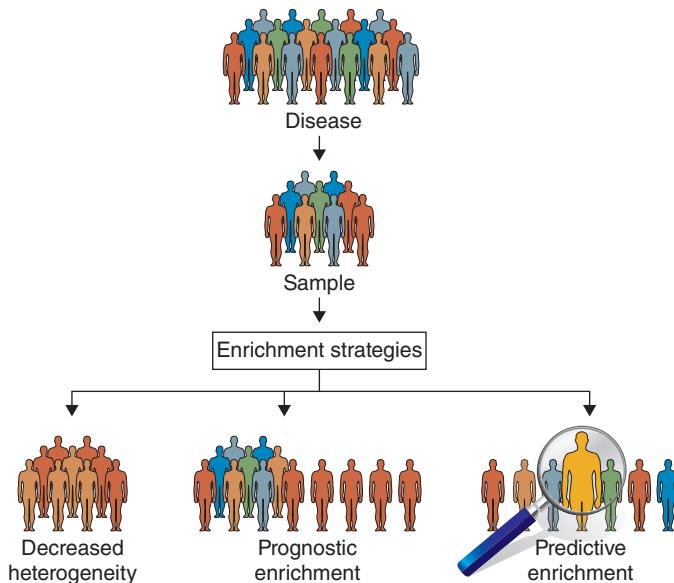


FIGURE 5-8 The basis for precision medicine. The notion of precision medicine evolved, in part, from clinical trial design. From the entire population of patients with the disease of interest, a sample cohort of individuals is enrolled in the trial that ideally is representative of the entire distribution. Enrichment strategies developed to decrease heterogeneity or increase the representation of individuals with a high risk of observed outcomes (prognostic enrichment) facilitate trial conduct but do not necessarily improve precision in defining treatment response. The predictive enrichment strategy utilizes both trial participant characteristics and data from experiments conducted before or during (adaptive design) the trial to improve the prediction of who is likely to have a more pronounced response to the treatment under study. (Reproduced with permission from EM Antman, J Loscalzo: Precision medicine in cardiology. *Nat Rev Cardiol* 13:591, 2016.)

to clinical trial design meant to improve therapeutic signal are illustrated. Decreasing heterogeneity and enriching the study population will enhance the effect size, but these strategies are based on analyses of prior data sets that define those individuals who are more likely than not to respond to a therapy. By contrast, the notion of predictive enrichment follows from the information provided by a detailed, big data-driven analysis of individuals that explores phenotypic and genomic features used to predict response. These features need not be precisely met by each patient; however, they can be collated or clustered to define a reasonably sized cohort predicted to respond in a particular way within certain confidence bounds. In this way, the boundaries to the practice of precision medicine are imprecise strictly speaking, but sufficiently predictive to be practical from the perspectives of clinical care and cost-effectiveness.

FURTHER READING

- ANTMAN EM, LOSCALZO J: Precision medicine in cardiology. *Nat Rev Cardiol* 13:591, 2016.
- CHENG F et al: Comprehensive characterization of protein-protein interactions perturbed by disease mutations. *Nat Genet* 53:342, 2021.
- CHENG F et al: A genome-wide positioning systems network algorithm for *in silico* drug repurposing. *Nat Commun* 10:3476, 2019.
- GREENE JA, LOSCALZO J: Putting the patient back together—Social medicine, network medicine, and the limits of reductionism. *N Engl J Med* 377:2493, 2017.
- KITSAK M et al: Tissue specificity of human disease module. *Sci Rep* 6:35241, 2016.
- LEE LY, LOSCALZO J: Network medicine in pathobiology. *Am J Pathol* 189:1311, 2019.
- LEOPOLD JA et al: The application of big data to cardiovascular disease: Paths to precision medicine. *J Clin Invest* 130:29, 2020.
- LOSCALZO J et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. *Mol Syst Biol* 3:124, 2007.
- MARON BA et al: Individualized interactomes for network-based precision medicine in hypertrophic cardiomyopathy with implications for other clinical pathophenotypes. *Nat Commun* 12:873, 2021.
- MENCHE J et al: Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347:1257601, 2015.
- SAMOKHIN AO et al: NEDD9 targets COL3A1 to promote endothelial fibrosis and pulmonary arterial hypertension. *Sci Transl Med* 10:eaap7294, 2018.

intervention requires a particularly high bar of evidence that testing and intervention are both practical and effective.

 Because population-based screening and prevention strategies must be extremely low risk to have an acceptable benefit-to-harm ratio, the ability to target individuals who are more likely to develop disease could enable the application of a wider set of potential approaches and increase efficiency. Currently, there are many types of data that can predict disease incidence in an asymptomatic individual. Germline genomic data have received the most attention to date, at least in part because mutations in high-penetrance genes have clear implications for preventive care (Chap. 467). Women with mutations in either *BRCA1* or *BRCA2*, the two major breast cancer susceptibility genes identified to date, have a markedly increased risk (five- to twentyfold) of breast and ovarian cancer. Screening and prevention recommendations include prophylactic oophorectomy and breast magnetic resonance imaging (MRI), both of which are considered to incur too much harm for women at average cancer risk. Some women with *BRCA* mutations opt for prophylactic mastectomy to dramatically reduce their breast cancer risk. Although the proportion of common disease explained by high-penetrance genes appears to be relatively small (5–10% of most diseases), mutations in rare, moderate-penetrance genes, and variants in low-penetrance genes, also contribute to the prediction of disease risk. Most recently, polygenic risk scores combining information about variants across hundreds of genes are being evaluated for identifying individuals at high risk of coronary heart disease and other conditions. The advent of affordable whole exome/whole genome sequencing is likely to speed the dissemination of these tests into clinical practice and may transform the delivery of preventive care.

Other forms of “omic” data also have the potential to provide important predictive information. Proteomics and metabolomics can provide insight into gene function, but it has proven challenging to develop reliable, predictive measures using these platforms. More recently, it has become possible to measure the presence of mutations in DNA circulating in the bloodstream and in stool, with early promising evidence that these assays can be used to detect cancer before existing screening tests.

In addition to “omic” data, imaging data are increasingly being integrated into risk-stratified prevention approaches as evidence grows about the predictive ability of these data. For example, coronary computed tomography (CT) scans are used in many preventive cardiology programs to inform decisions about beginning statin therapy when there is conflicting or uncertain information from other risk assessment approaches. Of course, these data may also be helpful in predicting the risk of harms from screening or prevention, such as the risk of a false-positive mammogram.

In addition to advances in risk prediction, there are several other reasons that screening and prevention are likely to gain importance in medical care in the near term. New imaging modalities are being developed that promise to detect changes at the cellular and subcellular levels, greatly increasing the probability that early detection improves outcomes. The rapidly growing understanding of the biologic pathways underlying initiation and progression of many common diseases has the potential to transform the development of preventive interventions, including chemoprevention. Furthermore, screening and prevention offer the promise of both improving health and sparing the costs of disease treatment, an issue that will continue to gain importance as long as health care costs in the United States remain a concern to patients, government agencies, and insurers.

This chapter will review the basic principles of screening and prevention in the primary care setting. Recommendations for specific disorders such as cardiovascular disease, diabetes, and cancer are provided in the chapters dedicated to those topics.

BASIC PRINCIPLES OF SCREENING

The basic principles of screening populations for disease were published by the World Health Organization in 1968 (Table 6-1).

In general, screening is most effective when applied to relatively common disorders that carry a large disease burden (Table 6-2). The five leading causes of mortality in the United States are heart diseases, malignant neoplasms, chronic obstructive pulmonary disease,

6

Screening and Prevention of Disease

Katrina A. Armstrong, Gary J. Martin

A primary goal of health care is to prevent disease or detect it early enough that intervention will be more effective. Tremendous progress has been made toward this goal over the past 50 years. Screening tests are available for many common diseases and encompass biochemical (e.g., cholesterol, glucose), physiologic (e.g., blood pressure, growth curves), radiologic (e.g., mammogram, bone densitometry), and cytologic (e.g., Pap smear) approaches. Effective preventive interventions have resulted in dramatic declines in mortality from many diseases, particularly infections. Preventive interventions include counseling about risk behaviors, vaccinations, medications, and, in some relatively uncommon settings, surgery. Preventive services (including screening tests, preventive interventions, and counseling) are different than other medical interventions because they are proactively administered to healthy individuals instead of in response to a symptom, sign, or diagnosis. Thus, the decision to recommend a screening test or preventive

TABLE 6-1 Principles of Screening

The condition should be an important health problem.
There should be a treatment for the condition.
Facilities for diagnosis and treatment should be available.
There should be a latent stage of the disease.
There should be a test or examination for the condition.
The test should be acceptable to the population.
The natural history of the disease should be adequately understood.
There should be an agreed policy on whom to treat.
The cost of finding a case should be balanced in relation to overall medical expenditure.

accidents, and cerebrovascular diseases. Thus, many screening strategies are targeted at these conditions. From a global health perspective, these conditions are priorities, but malaria, malnutrition, AIDS, tuberculosis, and violence also carry a heavy disease burden (**Chap. 472**).

Having an effective treatment for early disease has proven challenging for some common diseases. For example, although Alzheimer's disease is the sixth leading cause of death in the United States, there are no curative treatments and no evidence that early treatment improves outcomes. Lack of facilities for diagnosis and treatment is a particular challenge for developing countries and may change screening strategies, including the development of "see and treat" approaches such as those currently used for cervical cancer screening in some countries. A long latent or preclinical phase where early treatment increases the chance of cure is a hallmark of many cancers; for example, polypectomy prevents progression to colon cancer. Similarly, early identification of hypertension or hyperlipidemia allows therapeutic interventions that reduce the long-term risk of cardiovascular or cerebrovascular events. In contrast, lung cancer screening has historically proven more challenging because most tumors are not curable by the time they can be detected on a chest x-ray. However, the length of the preclinical phase also depends on the level of resolution of the screening test, and this situation changed with the development of chest CT. Low-dose chest CT scanning can detect tumors earlier and has been demonstrated to reduce lung cancer mortality by 20% in individuals who had at least a 30-pack-year history of smoking. The short interval between the ability to detect disease on a screening test and the development of incurable disease also contributes to the limited effectiveness of mammography screening in reducing deaths from some forms of breast cancer. At the other end of the spectrum, the early detection of prostate cancer may not lead to a difference in the mortality rate because the disease is often indolent and competing morbidities, such as coronary artery disease, may ultimately cause mortality (**Chap. 70**). This uncertainty about the natural history is also reflected in the controversy about treatment of prostate cancer, further contributing to the challenge of screening in this disease. Finally, screening programs can incur significant economic costs that must be considered in the context of the available resources and alternative strategies for improving health outcomes.

METHODS OF MEASURING HEALTH BENEFITS

Because screening and preventive interventions are recommended to asymptomatic individuals, they are held to a high standard for demonstrating a favorable risk-benefit ratio before implementation. In general, the principles of evidence-based medicine apply to demonstrating the efficacy of screening tests and preventive interventions, where randomized controlled trials (RCTs) with mortality outcomes are the gold standard. However, because RCTs are often not feasible,

observational studies, such as case-control designs, have been used to assess the effectiveness of some interventions such as colonoscopy for colorectal cancer screening. For some strategies, such as Pap smear screening for cervical cancer, the only data available are ecologic data demonstrating dramatic declines in mortality.

Irrespective of the study design used to assess the effectiveness of screening, it is critical that disease incidence or mortality is the primary endpoint rather than length of disease survival. This is important because lead time bias and length time bias can create the appearance of an improvement in disease survival from a screening test when there is no actual effect. Lead time bias occurs because screening identifies a case before it would have presented clinically, thereby creating the perception that a patient lived longer after diagnosis simply by moving the date of diagnosis earlier rather than the date of death later. Length time bias occurs because screening is more likely to identify slowly progressive disease than rapidly progressive disease. Thus, within a fixed period of time, a screened population will have a greater proportion of these slowly progressive cases and will appear to have better disease survival than an unscreened population.

A variety of endpoints are used to assess the potential gain from screening and preventive interventions.

1. *The absolute and relative impact of screening on disease incidence or mortality.* The absolute difference in disease incidence or mortality between a screened and nonscreened group allows the comparison of size of the benefit across preventive services. A meta-analysis of Swedish mammography trials (ages 40–70) found that ~1.2 fewer women per 1000 would die from breast cancer if they were screened over a 12-year period. By comparison, at least ~3 lives per 1000 would be saved from colon cancer in a population (aged 50–75) screened with annual fecal occult blood testing (FOBT) over a 13-year period, and an estimated 20–24 lives per 1000 would be saved over the entire 25-year period. Based on this analysis, colon cancer screening may actually save more women's lives than does mammography. However, the relative impact of FOBT (30% reduction in colon cancer death) is similar to the relative impact of mammography (14–32% reduction in breast cancer death), emphasizing the importance of both relative and absolute comparisons.
2. *The number of subjects screened to prevent disease or death in one individual.* The inverse of the absolute difference in mortality is the number of subjects who would need to be screened or receive a preventive intervention to prevent one death. For example, 731 women aged 65–69 would need to be screened by dual-energy x-ray absorptiometry (DEXA) (and treated appropriately) to prevent one hip fracture from osteoporosis.
3. *Increase in average life expectancy for a population.* Predicted increases in life expectancy for various screening and preventive interventions are listed in **Table 6-3**. It should be noted, however, that the increase in life expectancy is an average that applies to a population, not to an individual. In reality, the vast majority of the population does not derive any benefit from a screening test. A small subset of patients, however, will benefit greatly. For example, Pap smears do not benefit the 98% of women who never develop cancer of the cervix. However, for the 2% who would have developed cervical cancer, Pap smears may add as much as 25 years to their lives. Some studies suggest that a 1-month gain of life expectancy is a reasonable goal for a population-based screening or prevention strategy.

TABLE 6-3 Estimated Average Increase in Life Expectancy for a Population

SCREENING OR PREVENTIVE INTERVENTION	AVERAGE INCREASE
Mammography:	
Women, 40–50 years	0–5 days
Women, 50–70 years	1 month
Pap smears, age 18–65	2–3 months
Getting a 35-year-old smoker to quit	3–5 years
Beginning regular exercise for a 40-year-old man (30 min, 3 times a week)	9 months–2 years

^aAssuming an unscreened population.

ASSESSING THE HARMS OF SCREENING AND PREVENTION

Just as with most aspects of medical care, screening and preventive interventions also incur the possibility of adverse outcomes. These adverse outcomes include side effects from preventive medications and vaccinations, false-positive screening tests, overdiagnosis of disease from screening tests, anxiety, radiation exposure from some screening tests, and discomfort from some interventions and screening tests. The risk of side effects from preventive medications is analogous to the use of medications in therapeutic settings and is considered in the U.S. Food and Drug Administration (FDA) approval process. Side effects from currently recommended vaccinations are primarily limited to discomfort and minor immune reactions. However, the concern about associations between vaccinations and serious adverse outcomes continues to limit the acceptance of many vaccinations despite the lack of data supporting the causal nature of these associations.

The possibility of a false-positive test occurs with nearly all screening tests, although the definition of what constitutes a false-positive result often varies across settings. For some tests such as screening mammography and screening chest CT, a false-positive result occurs when an abnormality is identified that is not malignant, requiring either a biopsy diagnosis or short-term follow-up. For other tests such as Pap smears, a false-positive result occurs because the test identifies a wide range of potentially premalignant states, only a small percentage of which would ever progress to an invasive cancer. This risk is closely tied to the risk of overdiagnosis in which the screening test identifies disease that would not have presented clinically in the patient's lifetime. Assessing the degree of overdiagnosis from a screening test is very difficult given the need for long-term follow-up of an unscreened population to determine the true incidence of disease over time. Recent estimates suggest that as much as 15–40% of breast cancers identified

by mammography screening and 15–37% of prostate cancers identified by prostate-specific antigen testing may never have presented clinically. Screening tests also have the potential to create unwarranted anxiety, particularly in conjunction with false-positive findings. Although multiple studies have documented increased anxiety through the screening process, there are few data suggesting this anxiety has long-term adverse consequences, including subsequent screening behavior. Screening tests that involve radiation (e.g., mammography, chest CT) add to the cumulative radiation exposure for the screened individual. The absolute amount of radiation is very small from any of these tests, but the overall impact of repeated exposure from multiple sources is still being determined. Some preventive interventions (e.g., vaccinations) and screening tests (e.g., mammography) may lead to discomfort at the time of administration, but again, there is little evidence of long-term adverse consequences.

WEIGHING THE BENEFITS AND HARMS

The decision to implement a population-based screening and prevention strategy requires weighing the benefits and harms, including the economic impact of the strategy. The costs include not only the expense of the intervention but also time away from work, downstream costs from false-positive results, "incidentalomas" or adverse events, and other potential harms. Cost-effectiveness is typically assessed by calculating the cost per year of life saved, with adjustment for the quality of life impact of different interventions and disease states (i.e., quality-adjusted life-year). Typically, strategies that cost \$50,000–100,000 per quality-adjusted year of life saved are considered "cost-effective" (Chap. 4).

The U.S. Preventive Services Task Force (USPSTF) is an independent panel of experts in preventive care that provides evidence-based recommendations for screening and preventive strategies based on an assessment of the benefit-to-harm ratio (**Tables 6-4 and 6-5**). Because

TABLE 6-4 Screening Tests Recommended by the U.S. Preventive Services Task Force for Average-Risk Adults

DISEASE	TEST	POPULATION	FREQUENCY	CHAPTER
Abdominal aortic aneurysm	Ultrasound	Men 65–75 who have ever smoked	Once	
Alcohol misuse	Alcohol Use Disorders Identification Test	All adults	Unknown	453
Breast cancer	Mammography with or without clinical breast examination	Women 50–75	Every 2 years	
Cervical cancer	Pap smear	Women 21–65	Every 3 years	70
	Pap smear and/or HPV testing	Women 30–65	Every 5 years if HPV negative	
Chlamydia/gonorrhea	Nucleic acid amplification test on urine or cervical swab	Sexually active women <25	Unknown	189
Colorectal cancer	Fecal occult blood testing	45–75	Every year	70, 81
	Fecal immunochemical-DNA	45–75	Every 1–3 years	
	Sigmoidoscopy	45–75	Every 5 years	
	Colonoscopy (or occult blood testing combined with sigmoidoscopy)	45–75	Every 10 years	
Depression	Screening questions	All adults	Periodically	
Diabetes	Fasting blood glucose or HgbA1c	Adults overweight, obese, or with hypertension	Every 3 years	403
Hepatitis C	Anti-HCV antibody followed by confirmatory PCR	18–79	Once	
HIV	Reactive immunoassay or rapid HIV followed by confirmatory test	15–65	At least once	
Hyperlipidemia	Cholesterol	40–75	Unknown	407
Hypertension	Blood pressure	All adults	Periodically	277
Intimate partner violence	Screening questions	Women of childbearing age	Unknown	
Lung cancer	Low-dose computed tomography	Adults 50 to 80 years who have a 20 pack-year smoking history and currently smoke or have quit within the past 15 years	Yearly	
Obesity	Body mass index	All adults	Unknown	
Osteoporosis	DEXA	Women >65 or >60 with risk factors	Unknown	411

Abbreviations: DEXA, dual-energy x-ray absorptiometry; HCV, hepatitis C virus; HPV, human papillomavirus; PCR, polymerase chain reaction.

Source: Adapted from the U.S. Preventive Services Task Force 2017. www.uspreventiveservicestaskforce.org/Page/Name/uspstf-a-and-b-recommendations/.

TABLE 6-5 Preventive Interventions Recommended for Average-Risk Adults

INTERVENTION	DISEASE	POPULATION	FREQUENCY	CHAPTER
Adult immunization				
Tetanus-diphtheria		>18	Every 10 years	
Varicella		Susceptibles only, >18	Two doses	
Measles-mumps-rubella		Women, childbearing age	One dose	
Pneumococcal		>64	13 followed by 23 valent	
Influenza		>18	Yearly	
Human papillomavirus		Up to age 27	If not done prior	
Zoster		>60	Once	
Chemoprevention				
Aspirin	Cardiovascular disease	Aged 50–59 years with a ≥10% 10-year cardiovascular disease risk (bleeding risk may = benefit for some groups)		
Folic acid	Neural tube defects in baby	Women planning or capable of pregnancy		
Tamoxifen/raloxifene	Breast cancer	Women at high risk for breast cancer		
Vitamin D	Fracture/falls	>64 at increased risk for falls		

there are multiple advisory organizations providing recommendations for preventive services, the agreement among the organizations varies across the different services. For example, all advisory groups support screening for hyperlipidemia and colorectal cancer, whereas consensus is lower for breast cancer screening among women in their forties and for prostate cancer screening. Because the guidelines are only updated periodically, differences across advisory organizations may also reflect the data that were available when the guideline was issued.

For many screening tests and preventive interventions, the balance of benefits and harms may be uncertain for the average-risk population but more favorable for individuals at higher risk for disease. Although age is the most commonly used risk factor for determining screening and prevention recommendations, the USPSTF also recommends some screening tests in populations based upon the presence of other risk factors for the disease. In addition, being at increased risk for the disease often supports initiating screening at an earlier age than that recommended for the average-risk population. For example, when there is a significant family history of colon cancer, it is prudent to initiate screening 10 years before the age at which the youngest family member was diagnosed with cancer.

Although informed consent is important for all aspects of medical care, shared decision-making may be a particularly important approach to decisions about preventive services when the benefit-to-harm ratio is uncertain for a specific population. For example, many expert groups, including the American Cancer Society, recommend an individualized discussion about prostate cancer screening, because the decision-making process is complex and relies heavily on personal issues. Some men may decline screening, whereas others may be more willing to accept the risks of an early detection strategy. Recent analysis suggests that many men may be better off not screening for prostate cancer because watchful waiting was the preferred strategy when quality-adjusted life-years were considered. Another example of shared decision-making involves the choice of techniques for colon cancer screening (**Chap. 70**). In controlled studies, the use of annual FOBT reduces colon cancer deaths by 15–30%. Flexible sigmoidoscopy reduces colon cancer deaths by ~40–60%. Colonoscopy appears to offer a greater benefit than flexible sigmoidoscopy with a reduction in risk of ~70%, but its use incurs additional costs and risks. These screening procedures have not been compared directly in the same population, but models suggest that appropriate frequencies of each technique may be associated with similar numbers of lives saved and cost to society per life saved (\$10,000–25,000). Thus, although one patient may prefer the ease of preparation, less time disruption, and the lower risk of flexible sigmoidoscopy, others may prefer the sedation, thoroughness, and time interval of colonoscopy.

COUNSELING ON HEALTHY BEHAVIORS

In considering the impact of preventive services, it is important to recognize that tobacco and alcohol use, diet, and exercise constitute the

vast majority of factors that influence preventable deaths in developed countries. Perhaps the single greatest preventive health care measure is to help patients quit smoking (**Chap. 454**). However, efforts in these areas frequently require behavior changes (e.g., weight loss, exercise) or the management of addictive conditions (e.g., tobacco and alcohol use) that are often recalcitrant to intervention. Although these are challenging problems, evidence strongly supports the role of counseling by health care providers (**Table 6-6**) in effecting health behavior change. Educational campaigns, public policy changes, and community-based interventions have also proven to be important parts of a strategy for addressing these factors in some settings. Although the USPSTF found that the evidence was conclusive to recommend a relatively small set of counseling activities, counseling in areas such as physical activity and injury prevention (including seat belts and bicycle and motorcycle helmets) has become a routine part of primary care practice.

IMPLEMENTING DISEASE PREVENTION AND SCREENING

The implementation of disease prevention and screening strategies in practice is challenging. A number of techniques can assist physicians with the delivery of these services. An appropriately configured electronic health record can provide reminder systems that make it easier for physicians to track and meet guidelines. Some systems give patients secure access to their medical records, providing an additional means to enhance adherence to routine screening. Systems that provide nurses and other staff with standing orders are effective for immunizations. The USPSTF has developed flow sheets and electronic tools to assist clinicians (<https://www.uspreventiveservicestaskforce.org/uspstf/information-health-professionals>). Many of these tools use age categories to help guide implementation. Age-specific recommendations for screening and counseling are summarized in **Table 6-7**.

Many patients see a physician for ongoing care of chronic illnesses, and this visit provides an opportunity to include a “measure of prevention” for other health problems. For example, a patient seen for management of hypertension or diabetes can have breast cancer

TABLE 6-6 Preventive Counseling Recommended by the U.S. Preventive Services Task Force (USPSTF)

TOPIC	CHAPTER REFERENCE
Alcohol and drug use	453, 456, 457
Genetic counseling for <i>BRCA1/2</i> testing among women at increased risk for deleterious mutations	79, 467
Nutrition and diet	332, 333
Sexually transmitted infections	136, 202
Sun exposure	61
Tobacco use	454

TABLE 6-7 Age-Specific Causes of Mortality and Corresponding Preventive Options

AGE GROUP	LEADING CAUSES OF AGE-SPECIFIC MORTALITY	SCREENING PREVENTION INTERVENTIONS TO CONSIDER FOR EACH SPECIFIC POPULATION
15–24	1. Accident 2. Homicide 3. Suicide 4. Malignancy 5. Heart disease	<ul style="list-style-type: none"> Counseling on routine seat belt use, bicycle/motorcycle/ATV helmets (1) Counseling on diet and exercise (5) Discuss dangers of alcohol use while driving, swimming, boating (1) Assess and update vaccination status (tetanus, diphtheria, hepatitis B, MMR, rubella, varicella, meningitis, HPV) Ask about gun use and/or gun possession (2,3) Assess for substance abuse history including alcohol (2,3) Screen for domestic violence (2,3) Screen for depression and/or suicidal/homicidal ideation (2,3) Pap smear for cervical cancer screening after age 21 (4) Discuss skin, breast awareness, and testicular self-examinations (4) Recommend UV light avoidance and regular sunscreen use (4) Measurement of blood pressure, height, weight, and body mass index (5) Discuss health risks of tobacco use, consider emphasis on cosmetic and economic issues to improve quit rates for younger smokers (4,5) Chlamydia and gonorrhea screening and contraceptive counseling for sexually active females, discuss STD prevention Hepatitis B, and syphilis testing if there is high-risk sexual behavior(s) or any prior history of sexually transmitted disease Hepatitis C screening starting at age 18 to 79 HIV testing Continue annual influenza vaccination
25–44	1. Accident 2. Malignancy 3. Heart disease 4. Suicide 5. Homicide 6. HIV	<p><i>As above plus consider the following:</i></p> <ul style="list-style-type: none"> Readdress smoking status, encourage cessation at every visit (2,3) Obtain detailed family history of malignancies and begin early screening/prevention program if patient is at significant increased risk (2) Assess all cardiac risk factors (including screening for diabetes and hyperlipidemia) and consider primary prevention with aspirin for patients at >3% 5-year risk of a vascular event (3) and statin therapy for higher risk patients Assess for chronic alcohol abuse, risk factors for viral hepatitis, or other risks for development of chronic liver disease Consider individualized breast cancer screening with mammography at age 40 (2)
45–64	1. Malignancy 2. Heart disease 3. Accident 4. Diabetes mellitus 5. Cerebrovascular disease 6. Chronic lower respiratory disease 7. Chronic liver disease and cirrhosis 8. Suicide	<ul style="list-style-type: none"> Consider prostate cancer screen with annual PSA and digital rectal examination at age 50 (or possibly earlier in African Americans or patients with family history) (1) Begin colorectal cancer screening at age 45 or 50 with fecal occult blood testing, flexible sigmoidoscopy, or colonoscopy (1) Reassess and update vaccination status at age 50 and vaccinate all smokers against <i>Streptococcus pneumoniae</i> at age 50 (6) Consider screening for coronary disease in higher-risk patients (2,5) Zoster vaccination at age 60 Begin mammography screening by age 50 Lung cancer screening at age 50 to 80 years if a 20 pack-year smoking history and currently smoke or have quit within the past 15 years, yearly.
≥65	1. Heart disease 2. Malignancy 3. Cerebrovascular disease 4. Chronic lower respiratory disease 5. Alzheimer's disease 6. Influenza and pneumonia 7. Diabetes mellitus 8. Kidney disease 9. Accidents 10. Septicemia	<p><i>As above plus consider the following:</i></p> <ul style="list-style-type: none"> Readdress smoking status, encourage cessation at every visit (1,2,3,4) One-time ultrasound for AAA in men 65–75 who have ever smoked Consider pulmonary function testing for all long-term smokers to assess for development of chronic obstructive pulmonary disease (4,6) Screen all postmenopausal women (and all men with risk factors) for osteoporosis Continue annual influenza vaccination and vaccinate against <i>S. pneumoniae</i> at age 65 (4,6) Screen for visual and hearing problems, home safety issues, and elder abuse (9) Consider fall prevention exercise intervention if at higher risk (9)

Note: The numbers in parentheses refer to areas of risk in the mortality column affected by the specified intervention.

Abbreviations: AAA, abdominal aortic aneurysm; ATV, all-terrain vehicle; HPV, human papillomavirus; MMR, measles-mumps-rubella; PSA, prostate-specific antigen; STD, sexually transmitted disease; UV, ultraviolet.

screening incorporated into one visit and a discussion about colon cancer screening at the next visit. Other patients may respond more favorably to a clearly defined visit that addresses all relevant screening and prevention interventions. Because of age or comorbidities, it may be appropriate with some patients to abandon certain screening and

prevention activities, although there are fewer data about when to “sunset” these services. For many screening tests, the benefit of screening does not accrue until 5–10 years of follow-up, and there are generally few data to support continuing screening for most diseases past age 75. In addition, for patients with advanced diseases and limited life

expectancy, there is considerable benefit from shifting the focus from screening procedures to the conditions and interventions more likely to affect quality and length of life.

FURTHER READING

- BRETTHAUER M et al: America, we are confused: The updated U.S. Preventive Services Task Force recommendation on colorectal cancer screening. *Ann Intern Med* 166:139, 2017.
- HAYES JH et al: Observation versus initial treatment for men with localized, low-risk prostate cancer: A cost-effectiveness analysis. *Ann Intern Med* 158:853, 2013.
- HUGOSSON J et al: Mortality results from the Goteborg randomized population-based prostate-cancer screening trial. *Lancet Oncol* 11:725, 2010.
- OEFFINGER KC et al: Breast cancer screening for women at average risk 2015. Guideline update from the American Cancer Society. *JAMA* 314:1599, 2015.
- US PREVENTIVE SERVICES TASK FORCE: Screening for colorectal cancer. US Preventive Services Task Force recommendation statement. *JAMA* 315:2564, 2016.

7

Global Diversity of Health System Financing and Delivery

Richard B. Saltman

Health care systems are highly complex organizations, with many interdependent components. In developed countries, health systems have traditionally been classified by their type of financing—i.e., either predominantly tax-funded (such as the National Health Service in England and publicly operated regional care systems in the four European Nordic countries) or predominantly statutory social health insurance (SHI)-funded (such as in Germany, the Netherlands, and France). Over the past several decades, however, there has been structural convergence in the technical characteristics of both funding arrangements and in the associated delivery systems, making analytic observations about differences across national systems more difficult.

A second confounding factor has been that former Soviet Bloc countries in Central and Eastern Europe, including the Russian Federation, have, since 1991, replaced their former Soviet-style Semashko models (a top-down, national government-controlled funding and delivery structure with a parallel Communist Party administrative apparatus) with various hybrid arrangements built on national government-run SHI financing. Distinctions across developed country health systems, especially in Europe, have been further compressed by inadequate resources in many publicly funded systems in an era of rapid clinical and technological change, triggering increased private sector funding and provision.

In middle-income developing countries, institutional structures in the health sector typically reflect the country's preindependence administrative framework. Mexico, for example, has a Spanish-derived configuration with health insurance as part of social insurance for formally employed workers (via Instituto Mexicano del Seguro Social), supplemented by tax-funded health services (Seguro Popular) provided for those with informal employment and all other citizens, as well as a separate program (Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado) for public employees. Countries such as India and Egypt, reflecting British influence, have predominantly tax-funded and publicly operated health systems. China is an exception, with an internally generated system that is publicly funded and operated, although recent Communist Party policy has been to introduce SHI-based

insurance with individual medical savings accounts (patterned after Singapore), promote private insurance, and expand private hospitals.

In lower-income developing countries, health services are typically provided by tax-funded public institutions, often with considerable inadequacies and sometimes with substantial copayments. It is important to note that governmentally organized systems in nearly all developing countries, as well as in former Soviet bloc countries and, to a lesser degree, in tax-funded developed countries, are supplemented to varying extents by a mix of private and/or employer-paid insurers and providers.

This chapter focuses on the individual patient care system: on the financing and delivery of individual clinical and preventive services. The individual patient care system is composed of the financing and delivery of necessary services to prevent death or serious harm ("rule of rescue"); to maintain quality of life; and to manage, reduce, and/or prevent the burden of illness on individual patients. While the technical dimensions of most clinical services are similar across countries, their organizational, social, and economic characteristics range widely. Health systems in both developed and developing countries exhibit substantial differences, for example, in access to care; in the design and reliance on quality assurance and provider payment mechanisms; in the relationship of primary care to hospital services; in the coordination of health care with home care and nursing home services; in the design and use of provider management strategies; in the way physicians work and are paid; in the decision-making roles of politically elected officials and of national, regional, and municipal governments; and in participation of both citizens and patients. These wide-ranging institutional and organizational characteristics reflect differing country contexts (geographical, social, economic, and political), differences in national culture (consisting of prioritized norms and values), and substantial variation in how health sector institutions are structured.

FINANCING INDIVIDUAL PATIENT CARE SERVICES IN DEVELOPED COUNTRIES

Funding for individual care services in developed countries comes from the particular national mix among four possible sources of revenue: national, regional, and/or municipal taxes; mandatory SHI; private health insurance (including employer-paid insurances); and out-of-pocket payments. Most countries have one preponderant payer, which then defines its funding arrangements and serves to frame the structure of its delivery system as well.

Total Health Expenditures The Organization for Economic Co-operation and Development (OECD) data from 2017 (adjusted for purchasing power parities) show that total health care expenditures in developed countries vary across a considerable range, tied to health system structure as well as national history and culture (**Table 7-1**).

Per capita health expenditure figures provide a different, specific measurement of available funds in a country's health sector (**Table 7-2**).

Tax-Funded Systems In the United Kingdom, 79% of all health care funding was furnished through general tax revenues allocated by the national government in its annual budget process (all figures from OECD for 2017). In Sweden, all public taxes combined raised 83.7% of total health care spending. Sweden's 21 regional-level elected governments provide approximately 70% of that 83.7%, with the remaining 13.7% of total health spending raised by national and municipal taxes. In Canada, 71% of total health spending was raised by tax revenues, with 66% of that 71% coming from provincial or territorial taxes, while 5% came from national and local government taxes.

In most tax-funded countries, a segment of the population also has individual-, company-, or union-purchased private complementary and/or supplemental insurance coverage. In Sweden, 2019 estimates are that about 600,000 individuals have private complementary policies in a total population of 9 million. In Denmark, 50% of the population purchase supplemental insurance, while 30% have complementary insurance (often purchased by employers) that pays for private sector services enabling them to bypass public sector queues. In Finland, many middle-class families purchase separate private health insurance for their children to enable them to bypass long waiting times

TABLE 7-1 Developed Country Total Health Expenditure (% GDP)

TAX FUNDED IN WESTERN EUROPE	SHI FUNDED IN WESTERN EUROPE	CENTRAL EUROPEAN		DEVELOPED ASIAN		DEVELOPED NORTH AMERICAN	
Ireland	7.2%	Belgium	10.3%	Latvia	6.0%	Singapore	4.5%
Spain	8.9%	Netherlands	10.1%	Poland	6.5%	South Korea	7.6%
UK	9.6%	Germany	11.2%	Czech Republic	7.2%	Japan	10.9%
Finland	9.6%	Switzerland	12.3%	Slovenia	8.2%		
Denmark	10.1%						
Sweden	11.0%						

Abbreviations: GDP, gross domestic product; SHI, social health insurance; UK, United Kingdom.

Source: The Organization for Economic Co-operation and Development (OECD) data.

for primary and secondary pediatric health care services. More than 400,000 Finnish children (in a total population of 5 million) have privately purchased policies. In England in 2015, individual-, employer-, and union-purchased private complementary insurance covered an estimated 10.5% of the population, or about 6 million people. In Canada, individuals are not allowed by law to purchase private complementary insurance (except for Supreme Court-ordered insurance for three backlogged surgical procedures in Quebec Province—2005 Chaoulli decision); however, approximately 65% of the population have employer-, union-, or private group-purchased supplemental insurance for non-publicly covered services such as outpatient pharmaceutical prescriptions and home care.

Social Insurance-Funded Systems In Western Europe, SHI funds have traditionally been organized on a private not-for-profit basis, but with statutory responsibilities under national law. When former Soviet Bloc countries in Eastern Europe regained their independence in 1991, they returned to pre-World War II SHI models, but because there was no remaining organizational infrastructure, these post-1991 arrangements typically became a single SHI fund, run as an arm of the national government. In the United States, the Medicare social insurance system for citizens over age 65, enacted in 1965, is organized as a single fund tied to the national Social Security (public pension) Administration, an independent agency within the national government, with reimbursement arrangements supervised by the Centers for Medicare and Medicaid Services (CMS) inside the Department of Health and Human Services. Medicare covers inpatient hospital care plus limited post-hospital nursing home services (Medicare Part A). Supplemental private insurance policies are bought by covered individuals to help pay for outpatient physician visits (Medicare Part B) and for outpatient pharmaceuticals (Medicare Part D).

In Germany, 85% of the population is enrolled in one of 120 not-for-profit, monthly premium-based private SHI funds. This figure includes all individuals with annual incomes below 54,500 euros, who are required by law to join an SHI fund, as well as those with higher incomes who choose to enroll or remain. Eleven percent of the population—all having annual incomes above the mandatory SHI enrollment ceiling of 54,500 euros—have opted out of the SHI system to voluntarily enroll in claims-based private health insurance, whereas 4% of the citizenry is enrolled in sector-specific public programs such as the military. Since 2009, all SHI members pay a flat tax on gross monthly income as a contribution (8.2% in 2018, up to an upper income limit of 49,500 euros), which is transferred by their SHI fund to a national pool,

and then redistributed back to their chosen fund on an individual risk-adjusted basis. Employers send 7.3% of each employee's salary to the same national pool. Special arrangements exist for payments from self-employed, retired, and unemployed workers. Since 1995, there has been a separate mandatory social insurance fund for long-term care (LTC), with an annual premium of 1.95% of each adult's gross monthly income, split 50%–50% with their employer. Pensioners since 2004 are required to pay the full 1.95% from their pensions. Childless SHI enrollees pay a surcharge of 0.25% of monthly gross income. Overall, 78% of all health care expenditures in Germany were paid from public and/or mandatory private SHI sources.

In the Netherlands since 2006, all adult citizens pay a fixed premium (about 1453 euros in 2019) to their choice among 35 private health insurers (not-for-profit and for-profit), with four large insurance groups having over 1 million members each. In addition, employers pay 6.95% of salary below 51,400 euros for each employee into a national health insurance fund. Self-employed individuals pay 4.85% into the national fund for taxable income up to the same limit. Retired and unemployed individuals also make payments. In addition to the individual premiums paid to their choice of private insurance fund, payments from the national health insurance fund, adjusted by individual age, sex, and health characteristics, also are made to the individual's chosen insurer. The Netherlands has a separate mandatory social insurance fund for LTC (the ABWZ, since 2015 the WLZ, and now only for residential nursing home care) to which each employee pays 9.5% of taxable income beneath 33,600 euros every year. Self-employed, unemployed, and retired individuals also are required to pay premiums to the WLZ. Overall, including SHI revenues, public spending provided 87% of total health expenditures in 2014.

In Estonia, a former Soviet Republic that re-established an SHI system in 1991 upon regaining its independence, there is one national SHI fund that is an arm of the national government. This fund collects mandatory payments of 13% from salaried workers and 20% from self-employed individuals, covering both health care and retirement pensions. Overall, including SHI revenues, public spending accounted for 74.5% of total health expenditures in 2017.

Singapore, Japan, South Korea, and Taiwan have predominantly SHI systems of funding for individual care services. In these Asian countries (except Japan), there is one SHI fund that typically is operated as an arm of the national government.

In Singapore, starting in 1983, all employees up to age 50 have been required to place 20% of their income (employers add 16% more) into a personal health savings account to pay for direct health care costs,

TABLE 7-2 Developed Country Per Capita Health Expenditures

TAX FUNDED IN WESTERN EUROPE	SHI FUNDED IN WESTERN EUROPE	CENTRAL EUROPEAN		DEVELOPED ASIAN		DEVELOPED NORTH AMERICAN	
Spain	\$2738	Belgium	\$4149	Latvia	\$874	South Korea	\$2043
Italy	\$2738	Germany	\$4714	Poland	\$809	Singapore	\$4083
UK	\$3958	Netherlands	\$4742	Czech Republic	\$1321	Japan	\$4233
Denmark	\$5565	Switzerland	\$9835	Slovenia	\$1834		
Sweden	\$5710						

Abbreviations: SHI, social health insurance; UK, United Kingdom.

managed in their name by the Singapore government, called a Medisave account. Medisave accounts have a maximum amount, are tax-exempt, and receive interest payments (currently set at 4%). Consistent with a Confucian emphasis on family, the funds that accumulate in the Medisave account can be spent on health care for family members as well. If the accumulated funds are not spent on health care during the insured's life, they become part of the individual's personal estate and are distributed as a tax-free inheritance to his or her designated heirs. In addition, Singaporean citizens are also automatically enrolled into a second government-run health insurance plan called MediShield that pays for supplemental catastrophic, chronic, and long-term care. While citizens can opt out, 90% of citizens remain in the program. The Singapore government also operates a third, wholly tax-funded payer called Medifund that, with approval of a local neighborhood committee, will pay hospital costs for 3–4% of the population who are recognized as indigent. In part reflecting the high level of mandatory individual saving, public funding provided only 54.5% of total health expenditures in 2016.

In South Korea, a state-run SHI system was established in 1977, which in 1990 covered 30.9% of total health care costs. This percentage paid by the SHI system rose to 40.5% of total costs in 2017, with national tax revenue covering 16.9%, leaving out-of-pocket expenses at a relatively high 34.4% of total costs. Although there are legal ceilings on total out-of-pocket copayments for each 6-month period, over 70% of Korean adults purchase an additional private Voluntary Health Insurance policy to cover these additional direct expenditures. In 2000, three types of public SHI funds were merged into a single national state-run fund. As of 2018, 6% of an employee's salary must be paid as a social insurance contribution into this fund, with employees and employers each paying 50% of that amount. In 2008, an additional SHI fund was introduced to pay for LTC, operated by the main state-run SHI fund to reduce administrative costs. Contributions to the LTC fund are set at 6.55% of the individual's regular SHI contribution, coupled with 20% copayments for institutional care and 15% copayments for home care services.

The United States There is no single preponderant source of health care spending in the United States. The federal government's CMS reported that, for 2017, private health insurance covered 34% of total health expenditures, Medicare (mandatory SHI program for all citizens over 65) covered 20%, Medicaid (a joint federal-state welfare program for low-income citizens) covered 17%, and out-of-pocket paid 10%. Sources of funds for these programs were 28% from the federal government, 17% from state and local governments, 28% from private households, and 20% from private business (e.g., employers). The World Bank set public funding in the United States at 50.2% of total health expenditures in 2017.

In 2010, the passage of the Affordable Care Act (ACA) extended privately provided but heavily regulated and federally subsidized health insurance to many low- and middle-income uninsured individuals and families. Since the same act reduced the availability of existing individually purchased private health insurance, the total increase in the number of newly covered individuals was less than expected. Insurance premium increases for 2017 rose from 20% to over 100%, depending on the particular state, with additional increases in up-front deductible requirements, raising questions about the long-term sustainability of the ACA initiative. The recent Republican administration sought to repeal major financial and tax elements of the ACA and to replace existing subsidy arrangements with a system of refundable tax credits toward the establishment of individual health savings accounts and/or purchase of private health insurance on open cross-state markets (currently, private health insurance in the United States remains controlled at the separate 50-state level of government).

■ DELIVERING INDIVIDUAL PATIENT CARE SERVICES IN DEVELOPED COUNTRIES

Hospital Services In Europe, hospitals in both tax-funded and SHI-funded health systems are mostly publicly owned and operated by regional or municipal governments. In tax-funded health systems,

most hospital-based physicians are civil servants, employed on a negotiated salary basis (often by a physician labor union), and subject to most of the usual advantages and disadvantages of being a public sector employee. There are somewhat more private hospitals in SHI-funded health systems. However, most larger hospitals are public institutions operated by local governments, and most hospital physicians (with the notable exception of the Netherlands, where they are private contractors organized in private group practices) are, like those in tax-funded systems, public sector employees. In most tax-funded European countries (but not continental SHI-funded countries), few specialist physicians have office-based practices, and in both tax- and SHI-funded systems, office-based specialists do not have admitting privileges to publicly operated hospitals.

Most public hospitals in both tax-funded and SHI-funded health systems are single free-standing institutions that can be classified into three broad categories by complexity of patients admitted and number of specialties available: (1) district hospitals (four specialties: internal medicine, general surgery, obstetrics, and psychiatry); (2) regional hospitals (20 specialties); and (3) university hospitals (>40 specialties). In addition, many countries have a number of small, 15- to 20-bed, freestanding, private (typically for-profit) clinics. Recently, some tax-funded countries have begun to merge district and regional hospitals in an effort to improve the quality of care and create financial efficiencies (for example, Norway; planned for Denmark, also for Ireland; however, failed Parliamentary passage and brought down the coalition government, in Finland in 2019). Institutional mergers can be difficult to negotiate among publicly operated hospitals, due to the role that these large institutions play as important care providers and as large employers in smaller cities and towns, especially given political and union concerns about maintaining current employment levels. In the United States, financial and reimbursement pressures triggered by the implementation of the 2010 ACA have generated a number of private sector hospital mergers into larger hospital groups.

In tax-funded health systems, publicly funded patients who are admitted for an elective procedure cannot choose their specialist physician (except private-pay patients in "pay beds" in National Health Service [NHS] hospitals in England). Specialists are assigned by the clinic to a patient based on availability, with both junior and senior doctors placed in rotation.

Capital costs (buildings, large medical equipment) are publicly funded in all tax-funded systems and in most traditional SHI systems. For example, in Germany, capital costs for public hospitals are paid for by the regional governments. As a result, new capital investment is often allocated politically, according to location and political priorities. In Finland, local politicians in the 1980s would say that it "takes 10 years to build a hospital," meaning that it took that long to become a political priority for the regional government that controlled capital expenditures. Local politicians would therefore regularly overbuild when they got their one opportunity to obtain new capital.

Recently, efforts have been made to make public hospitals more responsible for their use of capital. In the Netherlands, public hospitals were shifted into private not-for-profit entities that are expected either to fund new capital from operating surplus or to borrow the funds from a bank based on a viable business plan. In England, more than 100 hospitals have been built using the Public Finance Initiative (PFI) program, in which private developers build turn-key facilities (thus taking capital costs off the public borrowing limit), and then rent these facilities back to the NHS and/or the relevant NHS Foundation Trust. In Sweden and Finland, while capital equipment is now a cost on hospital operating budgets, large new capital equipment and major building renovations remain politically driven processes often with extensive delays. In Stockholm County, the New Karolinska University Hospital opened in 2018 was built and is managed by a separate nonprofit public-private company.

In Singapore and South Korea, both of which are SHI funded, larger hospitals are publicly operated. However, there are a substantial number of smaller private clinics typically owned by specialist physicians. In the United States, the passage of the 2010 ACA has triggered the selling of many private specialist group practices to hospital groups,

transforming previously independent practicing physicians into hospital employees.

Primary Care Services Most primary health care in SHI-funded health systems, and also in an increasing number of tax-funded health systems (except in low-income areas of some large cities), is delivered by independent private general practitioners (GPs), working either individually or in small privately owned group practices. Recent changes in tax-funded health systems include Norway, where most primary care moved from municipally employed physicians to private-practice GPs in 2003, and Sweden, where, following a 2010 change in national reimbursement requirements, new privately owned not-for-profit and for-profit GP practices were established and now deliver 50% of all primary care visits.

In England, most primary care physicians are private GPs who are contractors to the NHS, working either independently or in small group practices. These private GPs own their own practices, which they can sell when they retire. However, as part of the original agreement to convince physicians to support the establishment of the NHS in 1948 (which most physicians strongly opposed), private GPs also receive a national government pension upon retirement. In the inner cities in England, there are some larger primary health clinics.

In 2001, England's private primary care doctors were organized into geographically based Primary Care Trusts (PCTs). These PCTs were allocated 80% of the total NHS budget to contract for elective hospital services required by their patients with both NHS hospital trusts as well as private hospitals. In 2013, PCTs were restructured into Clinical Commissioning Groups with similar contracting responsibilities.

In 2004, the Quality Outcomes Framework (QOF) was introduced as a quality of care-tied approach to providing additional income for NHS GPs. This regulatory mechanism in 2010 set 134 different standards for best practice primary care in four main domains: 86 clinical, 36 organizational, 4 preventive service, and 3 patient experience. GP income grew on average by 25% through the introduction of the QOF with general practices averaging 96% of possible QOF points. Total spending on QOF in 2014 in England consumed 15% of all primary care expenditures.

In April 2019, a slightly revised QOF contract was implemented, which retired 28 low-value indicators, introduced 15 new more clinically appropriate indicators, added two Quality Improvement modules, and added a new personalized care adjustment option. Funding was only changed marginally.

Access for individuals to primary care services is considered good in SHI-funded systems such as those in Germany and the Netherlands. One often-cited reason is that private office-based physicians (both GPs and specialists) in these countries are paid on a modified fee-for-service basis. In Germany, office-based physicians are paid on a quarterly basis by the Sickness funds, acting jointly at the Länder (regional) level through a point-based system. A national agreement between the physician association and the association of sick funds establishes points for each clinical act. Similarly, the association of sick funds (led in each of Germany's 16 Länder by the fund with the most subscribers in that region) establishes a fixed budget for all office-based physician services for all sick fund patients each 3-month period. Retrospectively at the end of each period, the total number of points is divided into the sick funds' fixed allocation for office-based physicians for that Länder for that quarter, establishing the value of a point for that quarter. Subsequently, each office-based physician's point total is multiplied by that quarterly point value, resulting in that physician's total payment from the statutory sick funds.

In contrast to SHI systems, seeing a primary care doctor in a number of tax-funded health systems has become increasingly difficult over the past decade. In Sweden, in 2005, a "care guarantee" was introduced that required its predominantly publicly operated health centers to see a patient within 7 days after calling for an appointment. In Finland, where public primary health care centers used to provide most primary care visits, delays in getting public health center appointments have pushed up to 40% of all visits into a parallel occupational health system, as well as to publicly employed primary care physicians working privately in the afternoons.

In England in 2019, access to GP services has been labeled a "crisis," aggravated by a 6% fall in the number of practicing GPs, leading to delays of up to 30 days for an appointment in urban areas like London. A 2019 report by the King's Fund found that only 1 in 20 trainee GPs planned to work full time. Also in 2019, the Nuffield Trust published a report suggesting that future planning for primary care services in England should assume a permanent shortage of GPs, requiring large numbers of new nurse practitioners and other auxiliary personnel. In Central European countries that were formerly within the Soviet Bloc, primary care provision had to be newly established after independence was regained in 1991, since first-line care in the former Semashko model was provided in specialist polyclinics. Primary care doctors rapidly emerged as almost entirely private for-profit GPs, working on contract from the national SHI fund (Estonia, Hungary, North Macedonia), from state-regulated private insurance companies (Czech Republic), or from regional/municipal public payers (Poland). Private GPs in most Central European countries now are paid on a per-visited basis. This arrangement was heavily influenced by the structure of primary care in Germany, where private office-based GPs are paid according to a point-system-tied framework.

In Asian countries such as Singapore, South Korea, and Japan, most primary care is provided by private for-profit GPs working independently or in small group practices. Private GPs are reimbursed at a set per-service fee by the national SHI fund(s). Access to primary care physicians is considered good.

Developed countries have varying policies regarding access to individual preventive services. Health systems in most countries provide vaccinations and mammography as part of funded health care services. In the United States, most insured individuals—and in Canada, most covered residents—automatically receive an annual physical exam including full blood profiles. In Norway and Denmark, adult physical exams are provided only upon special request by the individual, and in Sweden, adult physical exams are provided only to pregnant women. In Sweden, adults who wish to know their cholesterol or prostate-specific antigen (PSA) levels have begun to purchase blood tests out-of-pocket from private laboratories. In England in 2019, the NHS announced it would stop providing PSA screening tests for prostate cancer, even to men who requested one, similarly forcing concerned patients to purchase private laboratory testing.

Patients must make copayments to see a primary care doctor in some tax-funded health systems and in most SHI countries. In tax-funded systems, for example, Swedish patients are required to make a county-council-set copayment for each primary care visit up to a national-government-set annual ceiling, after which ambulatory visits (both primary and outpatient specialist) are not charged. Finland has a fixed copay for public health center visits, while Denmark's private GP visits do not have a copayment. In England, there is no copayment for GP visits.

In SHI health care systems in Europe and in Asia, patients usually are responsible for a copayment for both primary and office-based specialist care. To defray these charges (and to pay for other nonfunded services), a high percentage of citizens typically purchase additional supplemental health insurance. In France, where 95% of patients in 2015 purchased private supplemental insurance, patients paid directly the full fee for 65% of outpatient primary and specialist services, reimbursed subsequently by both their SHI fund and their supplemental insurance carrier for all payments (after deductibles), while for 35% of services (for low-income individuals and certain high-cost procedures), full agreed prices were paid directly to providers by SHI.

Access to Elective Specialist Care Approximately half of all European health care systems have a gatekeeping system that requires referrals from primary care physicians in order to book hospital specialist visits (for publicly paid visits). In most tax-funded health systems (although not in most SHI systems), there are substantial waiting times, typically several months or more, for elective specialist appointments as well as for high-tech diagnostic and treatment procedures. Waiting times can be particularly long for cancer and other elective surgical or high-demand services. In Sweden, government figures from the

summer of 2017 showed that, nationally, only 5–10% of prostate cancer operations were performed within 60 days after diagnosis.

In the English NHS, waiting lists for elective surgery in 2019 were often 6 months or longer. In August 2017, there were over 4,000,000 patients on NHS waiting lists. In January 2018, what administrators termed “a severe flu season,” during which hospital emergency rooms were overwhelmed with elderly patients requiring admission, led to a national-level NHS decision to cancel all elective operating room procedures in all hospitals in England (>50,000 procedures in 1200 hospitals) for the entire month of January, further lengthening waiting lists. Regarding quality of care, again in England, a March 2018 report from the national Office of Health Economics found that, in 2016 and 2017, up to three-quarters of patients who could have undergone key-hole procedures were forced to undergo open surgery, resulting in an estimated 1 million procedures each year that were more invasive than clinically necessary.

Delays in some tax-funded systems also are procedural. In England, for example, a patient who requires a further consultation with a second specialist typically has to return to their primary care physician for a second referral and then has to wait in the regular patient queue for that second appointment.

There is also substantial waiting time for radiologic imaging services in most tax-funded systems. In Malta, the tax-funded health system’s recent efforts to prioritize elective MRI investigations have succeeded in reducing waiting times from 18 months to 4 months. In both the Alberta and British Columbia Provinces in Canada, waiting times for a publicly funded nonemergency MRI can extend up to several months, whereas privately paid MRIs were available in both provinces within 1 week.

This issue of waiting times for specialist services in tax-funded health systems reflects a combination of growing demand (increasing/aging populations and changing clinical indications), financial constraints, and insufficient capacity, including inadequate physician working hours. For example, in the 1980s, when several surgical procedures for the elderly became more routine practice (e.g., hip replacement, coronary artery bypass graft, corneal lens implantation), the waiting list problem worsened. It had been mitigated somewhat through increased service capacity by the early 2000s, only to return as a growing policy challenge once public sector financial resources became constrained again after the 2008 global financial crisis. Timely cancer diagnosis and care continue to be a particularly sensitive issue, with tax-funded systems often taking several months for a patient to see an oncologist and then months more to begin treatment. In 2013 in Sweden, a newspaper journalist set off a political storm when he described women patients in one large county council (Malmö) who had to wait more than 40 days to receive the results from their breast cancer biopsy. In September 2019 in England, only 76.9% of patients with suspected cancer began treatment within 2 months of an urgent referral from a GP.

In response to pressure from national patient associations, a number of tax-funded health care systems introduced maximum waiting times for elective hospital procedures in the early 2000s. (Most Western European SHI systems do not have long waiting times or treatment guarantees for hospital care.) These maximum waiting times typically include initial primary care visits as well as specialist evaluations and treatment. In Denmark, a patient has the right to go to a different Danish public hospital for care after waiting 30 days without treatment. In Sweden, under the 2005 “waiting time guarantee,” an untreated patient’s local county council is required to pay for care in another county’s hospital after 180 days. In a parallel process at the European Union (EU) level, beginning in 1997, the EU Court of Justice steadily expanded the right of all EU citizens to travel to another EU country in order to receive “timely” care, with their home country health system required to pay for that care.

In private not-for-profit SHI-funded health systems such as in Germany and Switzerland, waiting times for specialist visits and hospital procedures are typically a few weeks to 1 month. In the SHI system in France, which is more centrally organized and funded (part of the Napoleonic tradition of public administration), ongoing disputes

about insufficient central government funding for public hospitals and staff salaries led in March 2019 to 9 months of hospital staff strikes, particularly in accident and emergency departments. In November 2019, the national government announced that it would take over 10 billion euros in public hospital debt as part of an effort to reverse staff cutbacks, bed and operating theater closures, and personnel flight to the private sector.

Long-Term Care Services LTC (consisting of residential and home-based services) consumes a relatively small but increasing proportion of gross domestic product (GDP) in developed countries. In 2016, Norway (2.95% GDP), Sweden (2.87% GDP), and the Netherlands (2.64% GDP) all spent more than one-fourth of their total health expenditures on LTC (Eurostat and OECD figures). More than one-fifth of all health care expenditures went to LTC in Belgium (2.16% GDP), Ireland (1.55% GDP), and Denmark (2.5% GDP). Lower-spending countries included the United Kingdom (18% of health expenditures; 1.75% GDP), Germany (12% of expenditures; 1.33% GDP), and Spain (9% of expenditures; 0.81% GDP). In the United States, official figures put total LTC expenditures in 2016 at 4.9% of total health expenditures, or 0.9% of total GDP. (Note that these figures do not include emergency, inpatient, or outpatient hospital costs generated by elderly patients.)

Since nursing home care is more expensive than home care (nursing home care requires the provision of housing, food, and around-the-clock care providers), government policymakers seek to keep the elderly and the chronically ill out of nursing homes for as long as feasible. Moreover, in developed countries like Sweden, Norway, and the United States, some 70% of all home care services come from informal caregivers: spouses, children (typically daughters), neighbors, and nonprofit community groups. While some SHI systems (e.g., Germany) have separate public LTC insurance (funded by mandatory premiums paid by all adults) that make available cash payments for LTC that can be used to compensate informal caregivers, most policymakers work hard to not monetize what is a large amount of essentially free care. Indeed, policymakers actively seek to encourage those providing these services to continue to do so as long as possible, trying to postpone caregiver burnout by providing support services such as free respite care, special call-in lines for caregiving advice, pension points toward retirement for the informal caregiver (Nordic countries), and free day-care center services.

In most tax-funded and SHI-funded European countries, home care services are organized at the municipal government level. In tax-funded systems, these services are also delivered mostly by municipal employees, working according to union-negotiated protocols. In some European SHI systems, and recently in tax-funded Sweden and Finland, private companies also provide home care services on contract to municipal governments. In combination with national legislation, these municipal systems also provide important support for informal caregivers, since the financial costs of caring for adults in their own home are substantially less than providing housing, food, and caregiver support in publicly funded homes for the aged or in nursing homes.

A high proportion of nursing homes in European tax-funded and SHI-funded health systems are publicly owned facilities operated by municipal governments; in some instances, in SHI-funded systems (Israel, the Netherlands), they are operated by private not-for-profit organizations. Recently, in some tax-funded systems (e.g., Sweden), private for-profit chains have begun to open nursing homes that are funded on a contract basis with local municipal governments. Costs for nursing home care can be expensive: in Norway, the cost per patient is often over \$100,000 per year in a publicly funded home, with the patient responsible for paying up to 80% depending on the family’s economic status. In Sweden, patients living in publicly funded nursing homes in Stockholm County pay a relatively small official fee, but they also pay room rent and up to 2706 Swedish krona (SEK) per month (about \$270 U.S. dollars [USD]) for food out of their monthly public pension payments.

In 2012, in an effort to reduce demand for expensive hospital and nursing home services, Norway and Denmark began elderly care reforms that shifted service delivery as well as funding responsibilities

to municipal governments. Among innovations in Norway, municipalities are required to establish a municipal acute bed unit (MAU) to treat stable elderly patients and provide observation beds for evaluation. Partial funding for these units is provided by the four public regional health care administrations. Some municipalities have also embedded primary care units inside their regional hospital to arrange discharge and to coordinate care for the chronically ill elderly. Norwegian municipalities are also responsible through their contracted (mostly private) primary care physicians to implement the National Pathways Program, which established treatment protocols for cross-sector conditions such as diabetes and cardiovascular conditions.

A differently configured structural innovation to better integrate LTC for the chronically ill elderly with clinical individual health services has been to consolidate both social and health care services within the same public administrative organization. In 2019, as part of health reforms in Ireland and Denmark and a proposed (unenacted) reform in Finland, as well as a pilot decentralization program in England for 2.8 million people in Greater Manchester, social and health care programs are to be administered by a single responsible agency.

In the SHI-funded system in the Netherlands, almost 7% of the population live in a residential home. National government legislation revised the structure of nursing home funding and care in 2015. Three acts restructured the separate public LTC SHI fund, which requires mandatory payments by 100% of Dutch adults, and introduced delivery-related reforms that reduced the number and overall cost of nursing home patients paid for by the fund. Determination of eligibility for public payment for nursing home care is now made by an independent national assessment body (the Centre for Needs Assessment). Moreover, municipal governments now play a stronger role in funding and delivering home care services. The reforms created social care teams that hold “kitchen table talks” to steer the elderly first toward seeking care from family, neighbors, churches, and other local community organizations before they qualify for publicly paid in-home care. In 2012, some 1.5 million people (12% of total population) provided informal care to ill or disabled persons, averaging 22 hours per week of care per person.

Home care recipients in the Netherlands can choose to set up a “personal budget,” using their public funding allocation to select their preferred individual care personnel (either publicly employed or publicly approved private providers). This arrangement also enables these home care recipients to determine the particular mix of services they want, as well as to augment the allocated public funds with personal funds. A number of innovative not-for-profit nursing homes have been created to provide additional services to elderly living in their neighborhood (primary care home visits), as well as terminal hospice care (e.g., the Saffier De Residentie Groep residences in The Haag).

In the United States, nursing home and home care are funded and delivered in a variety of different ways. For individuals who have minimal financial assets, nursing home costs are paid by a joint federal-regional (state) welfare program called Medicaid. Most state government Medicaid programs pay out more than 40% of their total budget for nursing home care. In the past, Medicaid did not pay for home care services. However, some states have programs with private for-profit and not-for-profit providers that provide home care as a way to forestall the need for the more expensive nursing home care.

Many private individuals take out private LTC insurance, typically from commercial insurance companies. These policies require individuals to make premium payments for years in advance (often 20 or more) before the individual learns whether they will, in fact, require home or nursing home care. Some private insurers have also raised premiums after individuals have paid in for many years and canceled policies if the new higher rate is not affordable. The 2010 ACA contained a new public LTC insurance program. However, the program was designed to be voluntary, and U.S. Department of Health and Human Services administrators decided in 2013 not to implement that portion of the law.

In addition to the tax-funded Medicaid program and privately purchased LTC insurance, many middle-class families pay for care from savings, by selling the elderly person’s home, or by direct contribution

from children and other family members. Expenses can reach between \$60,000 and \$100,000 per year depending on the location of a facility and who operates it.

Nursing home care in the United States is provided by a wide mix of private not-for-profit and for-profit providers, ranging from church-owned single-site homes to large stock market-listed companies. Many of these homes are purpose-built as assisted-living or memory-care facilities. Home care services are delivered by a mix of private not-for-profit and for-profit providers.

In Japan, a national LTC insurance fund was introduced in 2000. Although the new fund applies uniformly across the country, the program is administered by municipal governments and the premium level differs across municipalities, with an average monthly premium of 3000 yen (about \$30 USD). In South Korea, an SHI fund for LTC is funded by mandatory contributions of 4.78% of a person’s regular national health insurance contribution, with an additional 20% of total LTC expenditures provided by national government funds. The client copayment for home care is set at 15% of expenses and at 20% for residential care.

■ PHARMACEUTICALS

Pharmaceutical expenditures in developed countries (inpatient and outpatient combined) vary widely across different health system types, as well as between different countries within each institutional type. OECD figures for 2018 show drug expenditures in tax-funded countries in Western Europe ranging from 6.3% of total health expenditures (THE) in Denmark to 11.9% of THE in the United Kingdom and 18.6% of THE in Spain. In SHI-funded Western European systems, pharmaceuticals absorbed 7.5% of THE in the Netherlands, while in Germany, that figure was 14.1%. In the hybrid tax-funded SHI systems of Central Europe, the pharmaceutical percentage of THE is higher: 18.2% of THE in Estonia to 27.9% of THE in Hungary. Similarly, in Asian SHI systems, pharmaceuticals consumed 20.7% of THE in South Korea and 18.6% of THE in Japan. The OECD’s 2018 figures for pharmaceutical spending in North America are 12.0% of THE in the United States and 16.7% in Canada.

Contributing factors to this wide-ranging variation are (1) differences in national practice and prescription patterns reflecting differing cultural expectations; (2) the ratio problem (relatively fixed level of pharmaceutical costs due to international prices—the numerator—divided by a greatly varying per capita health expenditure cost in different developed country health systems); (3) the range and type of pharmaceutical price controls in each country; and (4) the degree of limitation placed on pharmaceutical supply, tied to formularies and/or explicit forms of drug rationing.

Most European health systems have tight national controls on the cost and, in some tax-based countries, on the availability of pharmaceuticals. Most European countries also use a number of different regulatory measures to limit prices and/or availability of both inpatient and outpatient drugs, including mandatory generic prescribing, reference pricing, patient copays (sometimes with an annual ceiling, after which copayments are no longer required), and (particularly in tax-funded systems) national formularies tied to clinical effectiveness. Norway, for example, allows only about 2300 different preparations—including dosage, delivery method, and box size—to be stocked by pharmacies. Prices for drugs can vary considerably across different European countries, tied to economic development and domestic pricing patterns. One consequence of these differential national pricing controls has been the development of a parallel import market, in which drug wholesalers and pharmacists in the more expensive countries purchase supplies from a cheaper market elsewhere in Europe.

Access to expensive drugs has also been intentionally limited in some tax-funded health systems in Europe. One basis for rationing has been rationing tied to quality-adjusted life-years (QALYs). Rationing also reflects a clash between strained public drug budgets and public pressure. For example, in the case of cancer drugs in England, the recommendation of the National Institute for Health and Care Excellence (NICE) against funding the breast cancer drug trastuzumab (Herceptin) was subsequently overturned by the Minister of Health.

Expensive cancer drugs continue to be rationed in England where the NHS Cancer Drug Fund, established in 2011 to provide access to non-NHS-provided drugs on a case-by-case basis, ran out of funds in 2015, forcing it to drop 25 of 83 covered drugs and close down for 3 months to restructure its operations.

As part of earlier medical patterns in Asian countries, office-based physicians traditionally filled prescriptions as well as prescribing drugs to patients. These sales also served to supplement their income in a setting of relatively low per-visit payments from state-run SHI funds. Concerned about cost and overuse, both Taiwan (in 1997, except for emergency cases or rural regions) and South Korea (for the whole country in 2003) implemented “separation reforms,” which ended these physician sales. In Japan, a series of fee and reimbursement reforms have trimmed the percentage of all prescriptions dispensed in 2016 by physicians to 26% of prescriptions filled.

■ GOVERNANCE AND REGULATION

Health care services in developed countries are steered, constrained, monitored, and (to varying degrees) assessed by governments and governmentally established and/or empowered bodies. Although these measures apply particularly to the financial efficiency of government-funded services, they also seek to promote patient and community safety, equity of access, and high-quality clinical outcomes. This oversight is often strongly focused on privately operated and contracted providers and insurers, although in principle, it applies to publicly operated organizations as well.

Governance consists of macro national-level policy, meso institutional-level management, and micro clinic-level care decisions. This complex mix of governance decisions is often shared among different national, regional, and local governments, depending on the degree of centralization, decentralization, or, recently, recentralization (e.g., Norway and Denmark). While most systems officially prioritize “good governance,” governance activities frequently comingle with political objectives as core policy concepts are developed and transformed into concrete organizational targets.

In Sweden, health system governance is shared among national, regional (county), and local municipal governments. The national government has responsibility to pass “frame” legislation, which establishes the basic structure of the system. To cite one example, until recently, the national government had limited an adult patient’s total copayments for outpatient physician care (specialist and primary care) and pharmaceuticals to 2800 SEK (about \$280 USD) for a 12-month period. The 20 regional governments, in turn, made policy decisions within that legislation, deciding how to apportion the specific copayments for each primary care and specialist outpatient visit. Since Swedes can self-refer to specialists, some counties double the copayment to hospital-based doctors to discourage unnecessary appointments. Similarly, fiscal policy normally is shared between the regional government, which raises about 70% of total health expenditures through its own county-set flat income tax, and the national government, which provides additional purpose-tied funds for national objectives such as consolidating open-heart surgery across county lines as well as supplementing lower tax receipts in rural counties with smaller working populations. However, this normal funding relationship across governments can change. In the early 1990s, the national government placed a “stop” on raising county taxes prior to Sweden’s admission in 1995 to the EU. In 2016, each of the 20 counties could set their own ceilings, which were almost all at 3300 SEK (about \$330 USD).

In Spain’s tax-funded health system (71.1% publicly funded in 2015), 17 regional “autonomous communities” were given full managerial responsibility for the provision of health services in a 1990s decentralization process, along with ownership of all publicly operated hospitals. The national government generates a substantial proportion of health care resources, which are included in the broad block grants it allocates to the regional governments, which then add regional tax revenue to make up the full public sector budget. In a mechanism to steer regional government operating policies in this decentralized environment, the national Spanish government established a joint federal-regional council to review quality and performance data (through the 2003 Health

System Cohesion and Quality Act). Italy’s tax-funded health system (75.8% publicly funded in 2014) similarly shares governance responsibilities between national and regional governments. Health services are provided by local health authorities (Azienda Sanitaria Locale) supervised by 20 regional governments within a nationally established governance framework, financed through a complicated mix of national and nationally stipulated but regionally collected taxes. Again, like Spain, the national government established a federal-regional government council, seeking to better coordinate care standards and information among the regions and with national government agencies. In 2006, the national government imposed strict financial plans on 10 regions that were systematically in deficit.

In Germany, where funding for its SHI-based health system is predominantly the responsibility of 120 private not-for-profit sickness funds, governance decisions are shared among these private sector sickness funds and public sector national, regional, and municipal governments. The sickness funds receive a risk-adjusted premium payment for each enrolled individual, according to a national government-determined formula, and from a national government-run health insurance pool. Most hospitals are owned and operated by municipal governments, while investment capital for structural renovations and new building comes from the 16 regional Länder taken from their tax revenues. Payment frameworks and amounts for public hospitals are negotiated between associations of these municipally owned hospitals and associations of the private sickness funds, without formal government participation.

Regulation is an essential element of an effective health care system and a key component of overall health system governance. Regulation incorporates both broad standard requirements that affect all organizations that operate in a country (e.g., hiring, firing, and wage decisions) as well as specific health sector-related regulations (e.g., proper handling, use, and disposal of low-grade nuclear waste from radiation treatments). Recent examples of health sector regulation in England, for example, include the following:

1. Requiring all cancer drugs adopted for use in the NHS to cost no more than \$41,268/QALY;
2. Requiring in their employment contract that junior doctors in hospitals work a specific number of Sundays; and
3. Requiring that all emergency department patients receive care within 4 h of their arrival.

A powerful tool that has the force of law, regulation can have substantial negative as well as positive effects. A well-known political science corollary of regulatory power is that “the right to regulate is also the right to destroy.” For example, in the United States, the federal Environmental Protection Agency, as part of its pursuit of cleaner air, issued wide-ranging regulatory orders setting performance standards that resulted in the closing of many West Virginia coal mines, with the loss of tens of millions of dollars of productive capacity and thousands of high-paying jobs, and likely contributing to social conditions that helped spawn that state’s high rates of opioid abuse among unemployed males. Similarly, in some tax-funded European systems, such as those in Sweden and England, there is growing pressure from public health advocates for national regulations to prohibit the making of a profit from publicly paid funds. In Sweden, the national government’s Reepalu report in 2016 honored a pledge made by the Social Democratic government to its Left (socialist) Party ally by calling for a legislated ban on profit-making in the provision of publicly funded health care services. The report’s publication triggered substantial divestment of existing investor-owned primary care, nursing home, and home care companies.

■ FUTURE CHALLENGES

Health systems in developed countries face continued challenges in the coming years. These include financial, organizational, and policy dilemmas for which institutionally viable, financially sustainable, and politically supportable solutions will be complicated to develop and difficult to implement. On the delivery side, a key question is whether

privately structured GP-based primary care is more efficient and effective than various clinic-based forms of primary care services. Recent movement in Northern and Central Europe toward more private GPs, along with continued private office-based primary care in much of Canada, the United States, and economically developed countries in Asia, raises complex policy issues for international organizations like the World Health Organization (WHO), as well as national policymakers. In the hospital sector, existing levels of clinical quality and patient responsiveness in publicly operated command-and-control institutions will increasingly have to compete with those of semi-autonomous public hospitals, as well as various types of private, sometimes very innovative providers. In the financing arena, continued pressure on publicly raised health system revenues is likely to erode longtime commitments in some tax-funded health systems to minimal patient copayments and low out-of-pocket funding.

An additional set of challenges will arise from recent commitments by international organizations like WHO to restructure health systems in developed countries to better address the social determinants of health. This new, incomplete strategy calls for a dramatic expansion of health sector responsibility to include a wide range of existing institutional arrangements in housing, education, work-life, and social and political decision-making. The influential 2010 Strategic Review of Health Inequalities in England entitled "Fair Society, Healthy Lives," led by Sir Michael Marmot, a British epidemiologist, called for the elimination of all "inequities in power, money, and resources." Separate from the political dimensions of this proposed new paradigm, how such fundamental societal change will be funded and implemented has yet to be addressed.

Looking forward, among the most essential challenges to national decision-makers in the coming period will be four specific health system imperatives:

1. Finding a more sustainable balance between ethics and funding.

Policymakers in publicly funded health systems face a growing gap between patient expectations of high-quality clinical care, staff expectations of better compensation, and the economic imperative of no new taxes. Recent research has suggested that SHI-funded health systems, faced with increasing aging and thus proportionally fewer employed, face a similar gap. While the present solidaristic foundation for raising collective revenues is insufficient, available nonsolidaristic tools (copayments, supplemental insurance, private pay) inevitably contribute to overall inequality. But what then are the realistic policy alternatives? The minimalist new policy goal necessarily will have to become one of raising new revenues while doing the least economic and social harm.

2. Developing better strategies to steer provider diversity.

Health systems in developed countries are becoming more diverse with more and different types of public owners: hospital trusts, state enterprises, and mixed public-private hospital owners/managers. There also are more and different types of private providers: not-for-profit community groups, foundations, and cooperatives, as well as for-profit small local entrepreneurs, large international companies, and risk capital funds (venture capital). Furthermore, new innovative delivery models are reorganizing traditional service boundaries: not-for-profit private nursing homes in the Netherlands also provide outpatient primary care to neighborhood elderly patients, as well as hospice care; Israeli technology companies combine high-tech home-based patient monitoring with standard medical and custodial home care services. Public pressure from citizens for more choice and better outcomes will pressure policymakers toward new, more accommodative health system arrangements. A 2019 national government report in Sweden on the hospital sector recommended a new emphasis on better access to out-of-office hours and out-of-hospital acute care by private as well as public providers.

3. Ensuring better coordination between social and health services.

Tax-funded and SHI-funded systems alike are under intense policy pressure to develop better strategies to integrate services for the chronically ill elderly, as a way to improve the quality of services that these patients receive and to keep them at home healthier and

longer, reducing expensive acute visits to hospitals and emergency departments. The clear delivery system goal will increasingly be to keep the elderly out of nursing homes and acute care facilities for as long as possible.

4. Building labor unions into provider innovation.

In many developed countries, health sector staff, including hospital physicians, are members of labor unions. Effective policymaking will require finding mechanisms to build these personnel unions into accelerated health system restructuring processes. This process will necessarily involve integrating unions into more innovative, flexible, fiscally sustainable organizational arrangements with contracts that reward active participation in organizational change, contracts that pay incentives to more productive employees, quicker reassignment and redundancy procedures (firing health sector workers can take a year or longer in some European health systems), and establishing profit-sharing payments to teams/unions, also in public sector organizations.

While the structure and complexity of resolving these specific organizational challenges will vary depending on a country's cultural and institutional context, the commonality of these problems suggests that health systems in the developed world require a new, broader range of targeted policy strategies and solutions.

■ FINANCING AND PROVIDING HEALTH SERVICES IN DEVELOPING COUNTRIES (See also Chap. 474)

Health systems in developing countries reflect a complex combination of the same core elements found in developed country systems (hospitals, primary care facilities, medical staff, pharmaceuticals) adapted to different, widely varying organizational, social, political, and economic contexts and conditions. System structure and provider institutions typically vary by differing national characteristics including historical relationships (Anglophone/Francophone/Hispanic/Soviet Semashko/American institutional and educational links); GDP and per capita annual national income (low- or middle-income developing countries); political norms and values; and ethnic and/or cultural mix. Predominantly public sector funding, particularly in lower-income countries, typically generates substantially lower levels of resources per capita than in developed countries and tends to be less reliable, particularly in countries where the economy is dependent on commodity exports.

Service delivery arrangements in developing countries, in turn, typically have higher provider-to-population ratios as well as, in public sector institutions, more mixed quality of care. In a number of middle-income developing countries, migration of trained medical staff to practice in higher-paying developed country health systems (often going to countries with historical relationships and/or where they received advanced training) further depletes available medical resources. In nearly all developing countries, private sector providers play an important supplemental role, with some middle-income developing countries like China currently encouraging their further development.

Most middle- and lower-income developing countries struggle to fund high-quality individual health services. Recent emphasis on universal health coverage has intensified that struggle. In middle-income developing countries (**Table 7-3**), World Bank data from 2016 show

TABLE 7-3 Middle-Income Developing Countries: Total Health Expenditure (% of gross domestic product)

Middle-Income Developing Countries

Kazakhstan	3.53%
Thailand	3.71%
Malaysia	3.80%
Turkey	4.31%
China	4.98%
Botswana	5.46%
Mexico	5.47%
Colombia	5.91%

TABLE 7-4 Low-Income Developing Countries: Total Health Expenditure (% of gross domestic product)

Low-Income Developing Countries	
Nigeria	3.65%
India	3.66%
Ethiopia	3.97%
Nepal	6.29%
Honduras	8.40%

a range of health expenditure rates as a percentage of GDP, including Kazakhstan at 3.53% of GDP, Thailand at 3.71%, Malaysia at 3.80%, Turkey at 4.31%, China at 4.98%, Botswana at 5.46%, Mexico at 5.47%, and Colombia at 5.91%. Total health spending in low-income developing countries (**Table 7-4**) ranges from 3.65% of GDP for Nigeria, 3.66% for India, 3.97% for Ethiopia, 6.29% for Nepal, to 8.40% for Honduras.

Given lower aggregate GDP levels, per capita annual expenditures are considerably less than those found in developed countries. In middle-income developing countries (**Table 7-5**), Thailand spent (2016 data in adjusted USD) \$221 annually per person, Kazakhstan spent \$262, Colombia spent \$340, Malaysia spent \$361, Botswana spent \$379, China spent \$398, Mexico spent \$461, and Turkey spent \$468. Among low-income developing countries (**Table 7-6**), Ethiopia spent \$27 per person annually, Nepal spent \$45, India spent \$62, and Nigeria spent \$79, whereas Honduras spent \$199.

China provides an interesting example of financing and service delivery development in middle-income developing countries. Financing reforms replaced fully publicly funded services with three new arrangements tied to work status and residence: (1) Urban Employee Basic Medical Insurance in 1998 (incorporating privately funded medical savings accounts—a concept pioneered in Singapore); (2) Urban Resident Basic Medical Insurance in 2007; and (3) New Rural Cooperative Medical Scheme in 2007. The urban employee program is an SHI model reflecting the rapid rate of economic growth and increasing incomes for urban workers. Starting in 2013, the Chinese government increasingly emphasized the development of new private hospitals and promotion of private insurance in urban areas. These and other health sector reforms became possible as continued strong economic growth over 30 years raised an estimated 300 million Chinese into the middle class, generating the requisite private as well as public revenues to underpin major structural health sector change.

Service delivery in developing countries varies widely in access, quality, and outcomes across and also within many developing countries. Medical services and tertiary institutions in urban areas of China, for example, operate at a substantially higher standard of service than those typically available in poorer rural regions. Similar disparities exist in wealthier parts of India such as Rajasthan, whereas in poorer states such as Bihar, primary care is mostly delivered by community “volunteers” with basic medical training, supervised by a GP.

Two critical challenges for all developing country health systems are contingent on generating adequate future funding flows. First, the current push from United Nations agencies to achieve universal health coverage will require additional public and private sector funding to

TABLE 7-5 Middle-Income Developing Countries: Per Capita Health Expenditures

Middle-Income Developing Countries	
Thailand	\$221
Kazakhstan	\$262
Colombia	\$340
Malaysia	\$361
Botswana	\$379
China	\$398
Mexico	\$461
Turkey	\$468

TABLE 7-6 Low-Income Developing Countries: Per Capital Health Expenditures

Low-Income Developing Countries	
Ethiopia	\$27
Nepal	\$45
India	\$62
Nigeria	\$79
Honduras	\$199

pay for the necessary new providers and services. Second, available funding will need to be more effectively targeted on needed and appropriate services, with minimized managerial inefficiencies and substantially less political corruption.

Both forms of expanded funding will be dependent on strong national and global economic growth, which in turn will require continued country-level economic and political reforms. Achieving both funding-related objectives will require considerable international as well as national effort.

FURTHER READING

BARBER SL et al: *Price Setting and Price Regulation in Health Care: Lessons for Advancing Universal Health Coverage*. Geneva, World Health Organization, Organization for Economic Co-operation and Development, 2019. <https://apps.who.int/iris/bitstream/handle/10665/325547/9789241515924-eng.pdf>.

FIGUERAS J, MCKEE M (eds): *Health Systems, Health, Wealth, and Societal Well-Being: Assessing the Case for Investing in Health Systems*. Maidenhead, Open University Press/McGraw-Hill Education, 2011. www.euro.who.int/_data/assets/pdf_file/0007/164383/e96159.pdf.

HASELTINE W: *Affordable Excellence: The Singapore Health Story*. Washington, Brookings Institution Press, 2013. www.brookings.edu/wp-content/uploads/2016/07/AffordableExcellencePDF.pdf.

KUHLMANN E et al (eds): *The Palgrave International Handbook on Healthcare Policy and Governance*. London, Palgrave MacMillan, 2015.

RICE T et al: *United States of America: Health System Review*. Health in Transition (HiT) Series 15 (3). Brussels, European Observatory on Health Systems and Policies, 2013. www.euro.who.int/_data/assets/pdf_file/0019/215155/HiT-United-States-of-America.pdf.

8

The Safety and Quality of Health Care

David W. Bates

Safety and quality are two of the central dimensions of health care. In recent years, it has become easier to measure safety and quality, and it is increasingly clear that performance in both dimensions could be much better. The public is—with good justification—demanding measurement and accountability, and payment for services will increasingly be based on performance in these areas. Thus, physicians must learn about these two domains, how they can be improved, and the relative strengths and limitations of the current ability to measure them.

Safety and quality are closely related but do not completely overlap. The Institute of Medicine has suggested in a seminal series of reports that safety is the first part of quality and that the health care system must first and foremost guarantee that it will deliver safe care, although quality is also pivotal. In the end, it is likely that more net clinical

benefit will be derived from improving quality than from improving safety, though both are important and safety is in many ways more tangible to the public. The first section of this chapter will address issues relating to the safety of care and the second will cover quality of care.

SAFETY IN HEALTH CARE

Safety Theory and Systems Theory *Safety theory* clearly points out that individuals make errors all the time. Think of driving home from the hospital: you intend to stop and pick up a quart of milk on the way home but find yourself entering your driveway without realizing how you got there. Everybody uses low-level, semiautomatic behavior for many activities in daily life; this kind of error is called a *slip*. Slips occur often during care delivery—e.g., when people intend to write an order but forget because they must complete another action first. *Mistakes*, by contrast, are errors of a higher level; they occur in new or nonstereotypic situations in which conscious decisions are being made. An example would be dosing of a medication with which a physician is not familiar. The strategies used to prevent slips and mistakes are often different.

Systems theory suggests that most accidents occur as the result of a series of small failures that happen to line up in an individual instance so that an accident can occur (Fig. 8-1). It also suggests that most individuals in an industry such as health care are trying to do the right thing (e.g., deliver safe care) and that most accidents thus result from defects in systems. Systems should be designed both to make errors less likely and to identify those that do inevitably occur.

Factors That Increase the Likelihood of Errors Many factors ubiquitous in health care systems can increase the likelihood of errors, including fatigue, stress, interruptions, complexity, and transitions. The effects of fatigue in other industries are clear, but its effects in health care have been more controversial until recently. For example, the accident rate among truck drivers increases dramatically if they work over a certain number of hours in a week, especially with prolonged shifts. A recent study of house officers in the intensive care unit demonstrated that they were about one-third more likely to make errors when they were on a 24-h shift than when they were on a schedule that allowed them to sleep 8 h the previous night. The American College of Graduate Medical Education has moved to address this issue by putting in place the 80-h workweek. Although this stipulation is a step forward, it does not address the most important cause of fatigue-related errors: extended-duty shifts. High levels of stress and heavy workloads also can increase error rates. Thus, in extremely high-pressure situations, such as cardiac arrests, errors are more likely to occur. Strategies such as using protocols in these settings can be helpful, as can simple recognition that the situation is stressful.

Interruptions also increase the likelihood of error and occur frequently in health care delivery. It is common to forget to complete an

action when one is interrupted partway through it by a page, for example. Approaches that may be helpful in this area include minimizing interruptions and setting up tools that help define the urgency of an interruption.

Complexity represents a key issue that contributes to errors. Providers are confronted by streams of data (e.g., laboratory tests and vital signs), many of which provide little useful information but some of which are important and require action or suggest a specific diagnosis. Tools that emphasize specific abnormalities or combinations of abnormalities may be helpful in this area.

Transitions between providers and settings are also common in health care, especially with the advent of the 80-h workweek, and generally represent points of vulnerability. Tools that provide structure in exchanging information—for example, when transferring care between providers—may be helpful.

The Frequency of Adverse Events in Health Care Most large studies focusing on the frequency and consequences of adverse events have been performed in the inpatient setting; some data are available for nursing homes, but much less information is available about the outpatient setting. The Harvard Medical Practice Study, one of the largest studies to address this issue, was performed with hospitalized patients in New York. The primary outcome was the adverse event: an injury caused by medical management rather than by the patient's underlying disease. In this study, an event either resulted in death or disability at discharge or prolonged the length of hospital stay by at least 2 days. Key findings were that the adverse event rate was 3.7% and that 58% of the adverse events were considered preventable. Although New York is not representative of the United States as a whole, the study was replicated later in Colorado and Utah, where the rates were essentially similar. Since then, other studies using analogous methodologies have been performed in various developed nations, and the rates of adverse events in these countries appear to be ~10%. Rates of safety issues appear to be even higher in developing and transitional countries; thus, this is clearly an issue of global proportions.

In the Harvard Medical Practice Study, adverse drug events (ADEs) were most common, accounting for 19% of all adverse events, and were followed in frequency by wound infections (14%) and technical complications (13%). Almost half of adverse events were associated with a surgical procedure. Among nonoperative events, 37% were ADEs, 15% were diagnostic mishaps, 14% were therapeutic mishaps, 13% were procedure-related mishaps, and 5% were falls.

ADEs have been studied more than any other error category. Studies focusing specifically on ADEs have found that they appear to be much more common than was suggested by the Harvard Medical Practice Study, although most other studies use more inclusive criteria. Detection approaches in the research setting include chart review and the use of a computerized ADE monitor, a tool that explores the database and identifies signals that suggest an ADE may have occurred. Studies that use multiple approaches find more ADEs than does any individual approach, and this discrepancy suggests that the true underlying rate in the population is higher than would be identified by a single approach. About 6–10% of patients admitted to U.S. hospitals experience an ADE.

Injuries caused by drugs are also common in the outpatient setting. One study found a rate of 21 ADEs per every 100 patients per year when patients were called to assess whether they had had a problem with one of their medications. The severity level was lower than in the inpatient setting, but approximately one-third of these ADEs were preventable.

The period immediately after a patient is discharged from the hospital appears to be very risky. A recent study of patients hospitalized on a medical service found an adverse event rate of 19%; about one-third of those events were preventable, and another one-third were ameliorable (i.e., they could have been made less severe). ADEs were the single leading error category.

Prevention Strategies Most work on strategies to prevent adverse events has targeted specific types of events in the inpatient setting, with nosocomial infections and ADEs having received the most attention. Nosocomial infection rates have been reduced greatly in intensive care

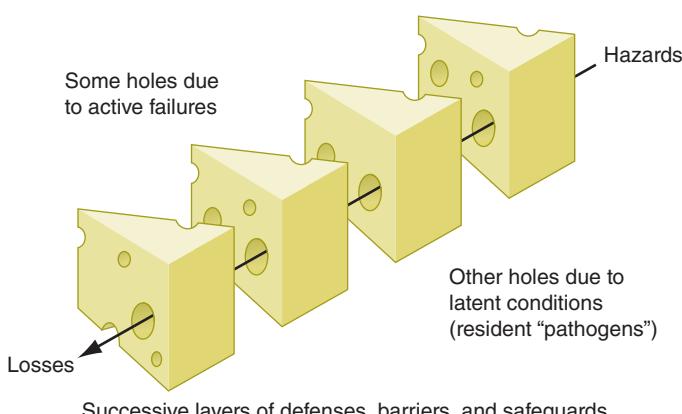


FIGURE 8-1 “Swiss cheese” diagram. Reason argues that most accidents occur when a series of “latent failures” are present in a system and happen to line up in a given instance, resulting in an accident. Examples of latent failures in the case of a fall might be that the unit is unusually busy and the floor happens to be wet. (Adapted from J Reason: BMJ 320:768, 2000.)

settings, especially by using checklists. For ADEs, several strategies have been found to reduce the medication error rate, although it has been harder to demonstrate that they reduce the ADE rate overall, and no studies with adequate power to show a clinically meaningful reduction have been published.

Implementation of checklists to ensure that specific actions are carried out has had a major impact on rates of catheter-associated blood-stream infection and ventilator-associated pneumonia, two of the most serious complications occurring in intensive care units. The checklist concept is based on the premise that several specific actions can reduce the frequency of these issues; when these actions are all taken for every patient, the result has been an extreme reduction in the frequency of the associated complication. These practices have been disseminated across wide areas in the state of Michigan.

Computerized physician order entry (CPOE) linked with clinical decision support reduces the rate of serious medication errors, defined as those that harm someone or have the potential to do so. In one study, CPOE, even with limited decision support, decreased the serious medication error rate by 55%. CPOE can prevent medication errors by suggesting a default dose, ensuring that all orders are complete (e.g., that they include dose, route, and frequency), and checking orders for allergies, drug-drug interactions, and drug-laboratory issues. In addition, clinical decision support can suggest the right dose for a patient, tailoring it to the level of renal function and age. In one study, patients with renal insufficiency received the appropriate dose only one-third of the time without decision support, whereas that fraction increased to approximately two-thirds with decision support; moreover, with such support, patients with renal insufficiency were discharged from the hospital half a day earlier. As of 2019, over 95% of U.S. hospitals had implemented CPOE, although the decision support often is still limited.

Another technology that can improve medication safety is bar coding linked with an electronic medication administration record. Bar coding can help ensure that the right patient gets the right medication at the right time. Electronic medication administration records can make it much easier to determine what medications a patient has received. Studies to assess the impact of bar coding on medication safety are under way, and the early results are promising. Another technology to improve medication safety is “smart pumps.” These pumps can be set according to which medication is being given and at what dose; the health care professional will receive a warning if too high a dose is about to be administered.

The National Safety Picture Several organizations, including the National Quality Forum and The Joint Commission, have made recommendations for improving safety. The National Quality Forum has released recommendations to U.S. hospitals about what practices will most improve the safety of care, and all hospitals are expected to implement these recommendations. Many of these practices arise frequently in routine care. One example is “readback,” the practice of recording all verbal orders and immediately reading them back to the physician to verify the accuracy of what was heard. Another is the consistent use of standard abbreviations and dose designations; some abbreviations and dose designations are particularly prone to error (e.g., 7U may be read as 70).

Measurement of Safety Measuring the safety of care is difficult and expensive, since adverse events are, fortunately, rare. Most hospitals rely on spontaneous reporting to identify errors and adverse events, but the sensitivity of this approach is very low, with only ~1 in 20 ADEs reported. Promising research techniques involve searching the electronic record for signals suggesting that an adverse event has occurred. These methods are not yet in wide use but will probably be used routinely in the future. Claims data have been used to identify the frequency of adverse events; this approach works much better for surgical care than for medical care and requires additional validation. The net result is that, except for a few specific types of events (e.g., falls and nosocomial infections), hospitals have little idea about the true frequency of safety issues.

Nonetheless, all providers have the responsibility to report problems with safety as they are identified. All hospitals have spontaneous reporting systems, and if providers report events as they occur, those events can serve as lessons for subsequent improvement.

Conclusions about Safety It is abundantly clear that the safety of health care can be improved substantially. As more areas are studied closely, more problems are identified. Much more is known about the epidemiology of safety in the inpatient setting than in outpatient settings. A number of effective strategies for improving inpatient safety have been identified and are increasingly being applied. Some effective strategies are also available for the outpatient setting. Transitions appear to be especially risky. The solutions to improving care often entail the consistent use of systematic techniques such as checklists and often involve leveraging of information technology. Nevertheless, solutions will also include many other domains, such as human factors techniques, team training, and a culture of safety.

■ QUALITY IN HEALTH CARE

Assessment of quality of care has remained somewhat elusive, although the tools for this purpose have increasingly improved. Selection of health care and measurement of its quality are components of a complex process.

Quality Theory Donabedian has suggested that quality of care can be categorized by type of measurement into structure, process, and outcome. *Structure* refers to whether a particular characteristic is applicable in a particular setting—e.g., whether a hospital has a catheterization laboratory or whether a clinic uses an electronic health record. *Process* refers to the way care is delivered; examples of process measures are whether a Pap smear was performed at the recommended interval or whether an aspirin was given to a patient with a suspected myocardial infarction. *Outcome* refers to what happens—e.g., the mortality rate in myocardial infarction. It is important to note that good structure and process do not always result in a good outcome. For instance, a patient may present with a suspected myocardial infarction to an institution with a catheterization laboratory and receive recommended care, including aspirin, but still die because of the infarction.

Quality theory also suggests that overall quality will be improved more in the aggregate if the performance level of all providers is raised rather than if a few poor performers are identified and punished. This view suggests that systems changes are especially likely to be helpful in improving quality, since large numbers of providers may be affected simultaneously.

The theory of *continuous quality improvement* suggests that organizations should be evaluating the care they deliver on an ongoing basis and continually making small changes to improve their individual processes. This approach can be very powerful if embraced over time.

Several specific tools have been developed to help improve process performance. One of the most important is the Plan-Do-Check-Act cycle (Fig. 8-2). This approach can be used for “rapid cycle” improvement of a process—e.g., the time that elapses between a diagnosis of pneumonia and administration of antibiotics to the patient. Some statistical tools, such as control charts, are often used in conjunction to determine whether progress is being made. Because most medical care includes one or many processes, this tool is especially important for improvement.

Factors Relating to Quality Many factors can decrease the level of quality, including stress to providers, high or low levels of production pressure, and poor systems. Stress can have an adverse effect on quality because it can lead providers to omit important steps, as can a high level of production pressure. Low levels of production pressure sometimes can result in worse quality, as providers may be bored or have little experience with a specific problem. Poor systems can have a tremendous impact on quality, and even extremely dedicated providers typically cannot achieve high levels of performance if they are operating within a poor system.

Data about the Current State of Quality A study published by the RAND Corporation in 2006 provided the most complete picture of

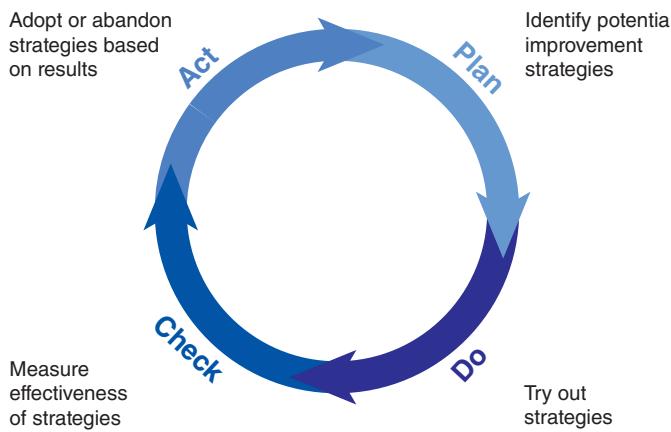


FIGURE 8-2 Plan-Do-Check-Act cycle. This approach can be used to improve a specific process rapidly. First, planning is undertaken, and several potential improvement strategies are identified. Next, these strategies are evaluated in small “tests of change.” “Checking” entails measuring whether the strategies have appeared to make a difference, and “acting” refers to acting on the results.

quality of care delivered in the United States to date. The results were sobering. The authors found that, across a wide range of quality parameters, patients in the United States received only 55% of recommended care overall; there was little variation by subtype, with scores of 54% for preventive care, 54% for acute care, and 56% for care of chronic conditions. The authors concluded that, in broad terms, the chances of getting high-quality care in the United States were little better than those of winning a coin flip.

Work from the Dartmouth Atlas of Health Care evaluating geographic variation in use and quality of care demonstrates that, despite large variations in utilization, there is no positive correlation between the two variables at the regional level. An array of data demonstrate, however, that providers with larger volumes for specific conditions, especially for surgical conditions, do have better outcomes.

Strategies for Improving Quality and Performance Many specific strategies can be used to improve quality at the individual level, including rationing, education, feedback, incentives, and penalties. *Rationing* has been effective in some specific areas, such as persuading physicians to prescribe within a formulary, but it generally has been resisted. *Education* is effective in the short run and is necessary for changing opinions, but its effect decays fairly rapidly with time. *Feedback* on performance can be given at either the group or the individual level. Feedback is most effective if it is individualized and is given in close temporal proximity to the original events. *Incentives* can be effective, and many believe that they will prove to be a key to improving quality, especially if pay-for-performance with sufficient incentives is broadly implemented (see below). *Penalties* produce provider resentment and are rarely used in health care.

Another set of strategies for improving quality involves changing the systems of care. An example would be introducing reminders about which specific actions need to be taken at a visit for a specific patient—a strategy that has been demonstrated to improve performance in certain situations, such as the delivery of preventive services. Another approach that has been effective is the development of “bundles” or groups of quality measures that can be implemented together with a high degree of fidelity. Many hospitals have implemented a bundle for ventilator-associated pneumonia in the intensive care unit that includes five measures (e.g., ensuring that the head of the bed is elevated). These hospitals have been able to improve performance substantially. Another technique is SCAMPs, or Standardized Clinical Assessment and Management Plans. These are care guidelines developed by clinicians who identify key steps in workflow and decisions to help improve the process outcomes.

Perhaps the most pressing need is to improve the quality of care for chronic diseases. The Chronic Care Model has been developed by Wagner and colleagues (Fig. 8-3); it suggests that a combination of

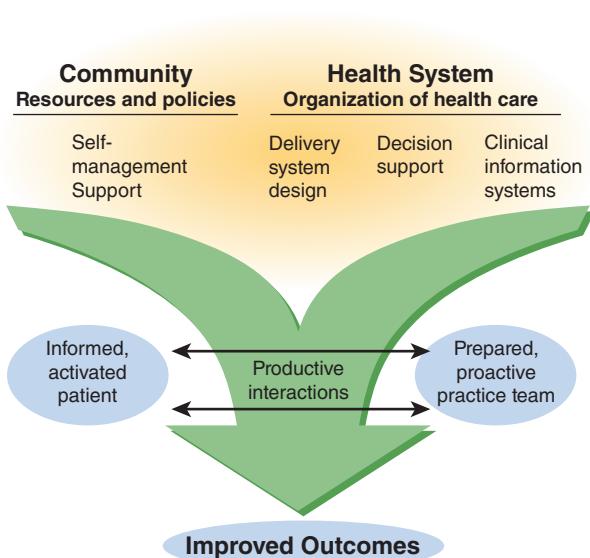


FIGURE 8-3 The Chronic Care Model, which focuses on improving care for chronic diseases, suggests that (1) delivery of high-quality care requires a range of strategies that must closely involve and engage the patient and (2) team care is essential. (From EH Wagner et al: Eff Clin Pract 1:2, 1998.)

strategies is necessary (including self-management support, changes in delivery system design, decision support, and information systems) and that these strategies must be delivered by a practice team composed of several providers, not just a physician.

Available evidence about the relative efficacy of strategies in reducing hemoglobin A_{1c} (HbA_{1c}) in outpatient diabetes care supports this general premise. It is especially notable that the outcome was the HbA_{1c} level, as it has generally been much more difficult to improve outcome measures than process measures (such as whether HbA_{1c} was measured). In this meta-analysis, a variety of strategies were effective, but the most effective ones were the use of team changes and the use of a case manager. When cost-effectiveness is considered in addition, it appears likely that an amalgam of strategies will be needed. However, the more expensive strategies, such as the use of case managers, probably will be implemented widely only if pay-for-performance takes hold.

The evidence linking better performance on quality metrics assessing process and outcomes varies greatly by condition. For example, there is strong evidence that performing Pap smears results in better outcomes in patients who develop cervical cancer, but the evidence for many other conditions is far more tenuous.

National State of Quality Measurement In the inpatient setting, quality measurement is now being performed by a very large proportion of hospitals for several conditions, including myocardial infarction, congestive heart failure, pneumonia, and surgical infection prevention; 20 measures are included in all. This is the result of the Hospital Quality Initiative, which represents a collaboration among many entities, including the Hospital Quality Alliance, The Joint Commission, the National Quality Forum, and the Agency for Healthcare Research and Quality. The data are housed at the Centers for Medicare and Medicaid Services, which publicly releases performance data on the measures on a website called *Hospital Compare* (www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospital-QualityInitits/HospitalCompare.html). These data are reported voluntarily and are available for a very high proportion of the nation's hospitals. Analyses demonstrate substantial regional variation in quality and important differences among hospitals. Analyses by The Joint Commission for similar indicators reveal that performance on measures by hospitals has improved over time and that, as might be hoped, lower performers have improved more than higher performers.

Public Reporting Overall, public reporting of quality data is becoming increasingly common. There are now commercial websites that have quality-related data for most regions of the United States, and

these data can be accessed for a fee. Similarly, national data for hospitals are available. The evidence to date indicates that patients have not made much use of such data, but that the data have had an important effect on provider and organization behavior. Instead, patients have relied on provider reputation to make choices, partly because little information was available until very recently and the information that was available was not necessarily presented in ways that were easy for patients to access. Problems still exist with quality metrics; many can be “gamed,” and even though providers are now nearly universally using electronic health records (EHRs), most metrics come from claims that include many inaccuracies. More metrics that leverage EHRs are sorely needed. However, many authorities think that, as more information about quality becomes available, it will become increasingly central to patients’ choices about where to access care.

Pay-for-Performance Currently, providers in the United States get paid the same amount for a specific service, regardless of the quality of care delivered. The pay-for-performance theory suggests that, if providers are paid more for higher-quality care, they will invest in strategies that enable them to deliver that care. The current key issues in the pay-for-performance debate relate to (1) how effective it is, (2) what levels of incentives are needed, and (3) what perverse consequences are produced. The evidence on effectiveness is limited, although a number of studies are ongoing. With respect to incentive levels, most quality-based performance incentives have accounted for merely 1–2% of total payment in the United States to date. In the United Kingdom, however, 40% of general practitioners’ salaries have been placed at risk according to performance across a wide array of parameters; this approach has been associated with substantial improvements in reported quality performance, although it is still unclear to what extent this change represents better performance versus better reporting. The potential for perverse consequences exists with any incentive scheme. One problem is that, if incentives are tied to outcomes, there may be a tendency to transfer the sickest patients to other providers and systems. Another concern is that providers will pay too much attention to quality measures with incentives and ignore the rest of the quality picture. The validity of these concerns remains to be determined. Nonetheless, it appears likely that, under health care reform, the use of various pay-for-performance schemes is likely to increase.

CONCLUSIONS

The safety and quality of care in the United States could be improved substantially. A number of available interventions have been shown to improve the safety of care and should be used more widely; others are undergoing evaluation or soon will be. Quality also could be dramatically better, and the science of quality improvement continues to mature. Implementation of value-based approaches such as accountable care that include pay-for-performance related to safety and quality should make it much easier for organizations to justify investments in improving safety and quality parameters, including health information technology. However, many improvements will also require changing the structure of care—e.g., moving to a more team-oriented approach and ensuring that patients are more involved in their own care. Payment reform focusing on value seems very likely to progress and will likely include both positive incentives and penalties related to safety and quality performance. Measures of safety are still relatively immature and could be made much more robust; it would be particularly useful if organizations had measures they could use in routine operations to assess safety at a reasonable cost, and substantial research is addressing this. Although the quality measures available are more robust than those for safety, they still cover a relatively small proportion of the entire domain of quality, and more measures need to be developed. The public and payers are demanding better information about safety and quality as well as better performance in these areas. The clear implication is that these domains will have to be addressed directly by providers.

FURTHER READING

BATES DW et al: Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 280:1311, 1998.

BATES DW et al: Two decades since to err is human: An assessment of progress and emerging priorities in patient safety. *Health Aff* (Millwood) 37:1736, 2018.

BERWICK DM: Era 3 for medicine and health care. *JAMA* 315:1329, 2016.

BRENNAN TA et al: Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med* 324:370, 1991.

CHERTOW GM et al: Guided medication dosing for inpatients with renal insufficiency. *JAMA* 286:2839, 2001.

INSTITUTE OF MEDICINE: Report: To err is human: Building a safer health system. 1999. <https://www.nap.edu/resource/9728/To-Err-is-Human-1999--report-brief.pdf>.

INSTITUTE OF MEDICINE: Crossing the quality chasm: A new health system for the 21st century. 2001. <https://www.nap.edu/catalog/10027/crossing-the-quality-chasm-a-new-health-system-for-the>.

LANDRIGAN C et al: Effect of reducing interns’ work hours on serious medical errors in intensive care units. *N Engl J Med* 351:1838, 2004.

MCGLYNN EA et al: The quality of health care delivered to adults in the United States. *N Engl J Med* 348:2635, 2003.

PRONOVO ST P et al: An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 355:2725, 2006. Erratum in: *N Engl J Med* 356:2660, 2007.

STARMER AJ et al: Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. *JAMA* 310:2262, 2013.

9

Diagnosis: Reducing Errors and Improving Quality

Gordon Schiff



Diagnosing patients’ illnesses is the essence of medicine. Patients present to doctors seeking an answer to the question, “What is wrong with me?” Ideally, no clinician would want to treat a patient without knowing the diagnosis or, worse yet, erroneously treat a misdiagnosed illness. From the earliest moments of medical school, the defining quest toward becoming a knowledgeable and proficient physician is learning how to put a diagnostic label on patients’ symptoms and physical findings, and clinicians pride themselves on being “good diagnosticians.” Yet the centuries-old paradigm of mastering a long list of diseases, understanding their pathophysiology, and knowing the cardinal ways they manifest themselves in signs and symptoms, while still of fundamental importance, is being challenged by new insights illuminated by the glaring spotlight of diagnostic errors. Basic internal medicine diseases, such as asthma, pulmonary embolism, congestive heart failure, seizures, strokes, ruptured aneurysms, depression, and cancer, are misdiagnosed at shockingly high rates, often with 20–50% of patients either being mislabeled as having these conditions (false-positive diagnoses) or having their diagnosis missed or delayed (false negatives). How and why do physicians so often get it wrong, and what can we do to both diagnose and treat the problem of delayed diagnosis or misdiagnosis?

Diagnosis is both an ancient art and a modern science. The current science of diagnosis, however, goes far beyond what typically comes to clinicians’ and patients’ minds when they conjure up images of state-of-the-art molecular, genetic, or imaging technologies. Improvements in diagnosis are just as likely to come from other areas, many with origins outside of medicine, as they are from advanced diagnostic testing modalities. These diverse sciences that the field of diagnostic safety has, and must, draw from include systems and human factors

engineering, reliability science, cognitive psychology, decision sciences, forensic science, clinical epidemiology, health services research, decision analysis, network medicine, learning health systems theory, medical sociology, team dynamics and communication, risk assessment and communication, information and knowledge management, and health information technology, especially artificial intelligence and clinical decision support. A clinician reading this chapter is likely to find this list of overlapping and intersecting domains quite daunting. However, rather than feeling overwhelmed, we urge readers to view them as the basic science supports that will ultimately make their lives easier and diagnosis more accurate and timely. Rather than feeling intimidated, clinicians should feel a sense of relief and assurance in understanding that good diagnosis does not rest entirely on their shoulders. Instead, it is a systems property, where an infrastructure and a team, one that especially includes the patient, can in a coordinated way work together to achieve more reliable and optimal diagnosis.

■ EMERGENCE OF DIAGNOSIS ERROR AS AN IMPORTANT PATIENT SAFETY ISSUE

Over the past decade, a series of studies culminating in a landmark report from the U.S. National Academy of Medicine (NAM), *Improving Diagnosis in Health Care*, have shone a spotlight on diagnostic errors. Reports from patient surveys, malpractice claims, and safety organizations, such as the ECRI and the National Patient Safety Foundation (now part of Institute for Healthcare Improvement), have found that diagnostic errors are the leading type of medical error. Although errors in diagnosis are defined in various ways, the NAM Committee defined diagnostic error as “the failure to (a) establish an accurate and timely explanation of the patient’s health problem(s) or (b) communicate that explanation to the patient.” One way to visualize diagnostic errors is through a Venn diagram (Fig. 9-1), which illustrates the fact that many things can go wrong in the diagnostic process (e.g., failure to ask an important history question, physical examination sign overlooked, laboratory specimen erroneously switched between two patients, x-ray not followed up), but this usually does not result in a wrong diagnosis or patient harm. Similarly, a patient can be misdiagnosed but unharmed, without any identifiable error in the care received. Our greatest concern is where these three circles intersect, with conservative estimates suggesting that 40,000–80,000 patients die each year in U.S. hospitals alone from diagnostic errors. The NAM report outlined eight recommendations that are the foundation for this chapter (Table 9-1).

■ NEW WAYS TO THINK ABOUT DIAGNOSIS AND DIAGNOSTIC ERRORS

Medical textbooks have historically given attention to “clinician reasoning” and associated cognitive heuristics and biases. Errors in clinical reasoning can be summarized in three broad groups: (1) hasty judgments, (2) biased judgments, and (3) inaccurate probability estimates. Research from cognitive psychology has identified scores of common mental shortcuts or “heuristics” humans are prone to use in everyday life, many of which are useful for efficient diagnosis but

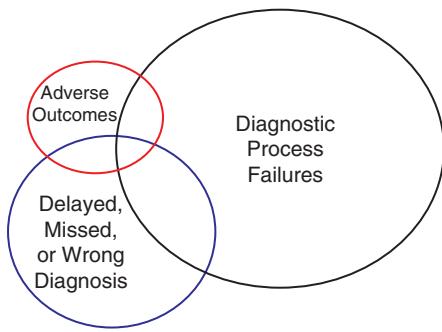


FIGURE 9-1 What is a diagnosis error? (Adapted from GD Schiff et al: *Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project*, in *Advances in Patient Safety: from Research to Implementation*. Vol. 2 Concepts and Methodology. Rockville, MD, 2005, pp. 255–278, and GD Schiff, L Leape: *Acad Med* 87:135, 2012.)

TABLE 9-1 National Academy of Medicine Recommendations for Improving Diagnosis in Health Care

1. Facilitate more effective teamwork in the diagnostic process among health care professionals, patients, and their families.
2. Enhance professional education and training in the diagnostic process in areas such as clinical reasoning; teamwork; communication with patients, families, and other health care professionals; and appropriate use of diagnostic tests.
3. Ensure that health information technologies support patients and health care professionals in the diagnostic process.
4. Develop and deploy approaches to identify, learn from, and reduce diagnostic errors and near misses in clinical practice including providing systematic feedback on diagnostic performance.
5. Establish a work system and culture that supports the diagnostic process and improvements in diagnostic performance.
6. Develop a reporting environment and medical liability system that facilitates improved diagnosis by learning from diagnostic errors and near misses.
7. Design a payment and care delivery environment that supports the diagnostic process.
8. Provide dedicated funding for research on the diagnostic process and diagnostic errors.

can also lead to biases and errors. Table 9-2 lists some of the common cognitive biases that can lead diagnosis astray (this topic is discussed further in Chap. 4).

However, clinicians will also benefit from having a better understanding of diagnosis as a “system” rather than just what takes place in clinicians’ minds. Classic teaching exhorting trainees and practicing physicians to have a broad differential and “high index of suspicion” for various diseases is challenged not only by these unconscious biases but also by limitations of human memory, information shortfalls, constrained encounter time, system process failures, and the myriad nonspecific symptoms that patients bring to clinicians. Many symptoms are self-limited, defy a precise diagnosis or etiology, and do not portend harmful outcomes. Insights from safety and cognitive sciences call for rethinking traditional approaches to diagnosis and suggest new approaches to overcome current limitations (Table 9-3).

■ UNCERTAINTY IN DIAGNOSIS

Given variations and overlap in ways patients present, illnesses evolve, and tests perform, it is often not possible or practical to “make” a definitive diagnosis, particularly in the primary care setting early in the course of a patient’s illness. Clinicians need to harness these uncertainties to both engineer situational awareness of where things

TABLE 9-2 Selected Cognitive Biases Contributing to Diagnostic Errors

1. Premature closure: accepting a diagnosis before it has been fully verified
2. Anchoring: tendency to fixate on a specific symptom or piece of information early in the diagnostic process with subsequent failure to appropriately adjust
3. Confirmation bias: tendency to look for confirming evidence to support one’s diagnostic hypothesis, rather than disconfirming evidence to refute it
4. Search satisfying: tendency to call off a search, satisfied once a piece of data or presumed explanation is found, and not considering/searching for additional findings or diagnoses
5. Availability bias: tendency to give too much weight to diagnoses that come more readily to mind (e.g., recent dramatic case)
6. Base-rate neglect: failing to adequately take into account prevalence of a particular disease (e.g., erroneously interpreting a positive test as indicating disease in a low-prevalence population using a test with 5% false-positive rate)
7. Knowledge deficit (on part of provider, with accompanying lack of awareness)
8. Framing bias: Judgement overly influenced by the way the problem was presented (how it was framed in words, settings, situations)
9. Social/demographic/stereotype bias: biases from personal or cultural beliefs about women, minorities, or other patient groups for whom prejudices may distort diagnostic assessment

TABLE 9-3 New Models for Conceptualizing Diagnosis and Diagnosis Improvement

TRADITIONAL WAYS OF THINKING ABOUT DIAGNOSIS AND DIAGNOSTIC ERROR	NEW PARADIGMS/BETTER WAYS TO THINK ABOUT DIAGNOSIS AND IMPROVING DIAGNOSIS
General	
A good diagnostician gets it right the first time, almost all of the time	Diagnosis is an inexact science with inherent uncertainties Goal is to minimize errors and delays via more reliable systems and follow-up
Lore of masterful/skillful academic expert diagnostician who knows/recalls everything; need to look to them if seeking diagnostic excellence	Less reliance on (fallible) human memory Quality diagnosis is based on well-coordinated distributed network/team of people and reliable processes All patients entitled to receive quality diagnosis, regardless of where and from whom they receive care
Diagnosis is the doctor's job	Co-production of diagnosis among clinicians (including lab, radiology, specialists, nurses, social workers) and, especially, the patient and family
Patients often viewed as overly anxious, exaggerating, time-consuming, questioning, with sometimes unreasonable demands and expectations	Patients are key allies in diagnosis; hold key information Need to address understandable/legitimate fears, desires for explanations Leveraging patient questions and questioning of diagnosis to stimulate rethinking the diagnosis where needed
Diagnosis and treatment as separate stages in patient care (i.e., make a diagnosis, then treat)	Prioritizing diagnostic efforts to target treatable conditions More integrated strategies and timing for testing and treatment depending on urgency for treatment
Clinical practices	
Order lots of tests to avoid missing diagnoses	Judicious ordering: targeted, well-organized data and testing Appreciation of test limitations (false positive or negative, incidental findings, overdiagnosis, test risks) and resulting harms
More referrals to avoid missing rarer/specialized diagnoses; concomitant utilization barriers (copays, prior authorization) to minimize overuse	"Pull systems" to lower barriers and make it easier to pose questions, obtain real-time virtual consults Co-management approaches to enable collaborative watch-and-wait conservative strategies where appropriate
Frequent empirical drug trials when uncertain of diagnosis	Conservative use of drugs to avoid confusing clinical picture or labeling patients with diseases they may not have
Physician attention/efforts to ensure disease screening	Automating, delegating clerical functions; teamwork to free up physician cognitive time
Diagnosis errors and challenges	
Diagnostic error viewed as a personal failing Errors classified as either "system" or "cognitive"	Many errors/delays rooted in processes and system design failures Errors multifactorial with interwoven, interacting, and inseparable cognitive and system factors
Errors are infrequent; hit-and-miss ways to learn about errors	Errors are common; systematic proactive follow-up is needed to recognize potential for errors Surveilling of high-risk situations and one's own diagnostic performance and outcomes
Clinicians' reactions: denial, defensive, others to blame, pointing to others also making similar errors	Culture of actively and nondefensively seeking to uncover, dig deep to learn from, and share errors and lessons
Dreading complex, frustrating diagnostic dilemmas	Welcoming/enjoying intellectual/professional challenges Adequate support (time, help, consultations) for more complex patients
Diagnoses as distinct labels, events	Diagnoses can be indistinct, interacting comorbidities, socially constructed, multifactorial, evolving over time, or have overlapping genotype-phenotype expressions
Documentation/communication	
Viewed as time-consuming, mindless, primarily to document for billing code and/or bulwark against malpractice claims	Documentation as useful tool for reflecting, crafting, sharing assessments, differential diagnosis, reflecting about unanswered questions Opportunities for decision support interacting with computer Notes open for patients to read to help understand and critique diagnosis
Say and write as little as possible about uncertainties, lest it be used against you in malpractice allegation	Share uncertainties to maximize communication and engagement with other caregivers, patients
Don't let patient know about errors so they don't become angry, mistrustful, or sue	Patients have right to honest disclosure; often find out about errors anyway (e.g., cancer evolves); anticipate, engage their concerns
Patients advised to call if not better; no news is good news (test results: "We'll call if anything is abnormal.")	Systematic proactive follow-up to close loop on tests and symptoms, to check how patient is doing, monitor outcomes
Global remedies	
Knowing/memorizing more medical knowledge	Knowing more about the patient (including psychosocial, past history, environmental contexts)
Attention to the "objective" data (physical exam, tests) to reliably make diagnoses	Renewed emphasis on history, history-taking, listening Acknowledgement of ubiquitous subjective cognitive biases; efforts to anticipate, recognize, counteract
Exhortations to have "high index of suspicion" of various diagnoses	Less reliance on memory recall of lectures/reading; more just-in-time info look-up Affordances, alerts to red flags engineered into workflow Delineation of "don't miss" diagnoses with design of context-relevant decision support reminders
Ensuring physician is copied on everything, thorough/voluminous notes, widespread reminders/alerts	Biggest problem is no longer lack of access to information, but rather information overload; strategies to organize, minimize
Continuing medical education (CME) courses to expand medical knowledge	Real-time, context-aware reminders of pitfalls, critical differential diagnoses, and key differentiating features. Ready access to medical references, second opinions

(Continued)

TABLE 9-3 New Models for Conceptualizing Diagnosis and Diagnosis Improvement (Continued)

TRADITIONAL WAYS OF THINKING ABOUT DIAGNOSIS AND DIAGNOSTIC ERROR	NEW PARADIGMS/BETTER WAYS TO THINK ABOUT DIAGNOSIS AND IMPROVING DIAGNOSIS
Redundancies, double-checks	Recognition that single, highly reliable systems are often better than multiple halfway solutions. Clear delineation of responsibilities for follow-up tasks
Fear of malpractice suits to motivate physicians to be more careful and practice defensive medicine	Drive out fear, making it safe to learn from and share errors Shared situational awareness of where pitfalls lurk
More accountability, financial incentives, and penalties tied to performance metrics	Clinician engagement in improvement based on trust, collaboration, professionalism, financial neutrality Metric modesty, recognizing many best practices yet to be defined/proven
More rules, requirements; target outliers for better compliance	Standardization with flexibility; learning from deviations
More time with patients	Better time spent with patients: offloading distractions, more efficient history collection/organization, longitudinal continuity, and, where needed, additional time to talk/think/explain during, before, or after visits Easier access for patients to reach or be seen by clinicians when experiencing symptoms
Reflex changes in response to errors	Avoiding “tampering,” which entails understanding/diagnosing difference between “special cause” versus “common cause” (random) variation

Source: Modified from GD Schiff: Quality and Safety in Health Care 2013.

can go wrong and create safety nets to protect patients against harms from delayed diagnosis and misdiagnosis. Terms such as *preliminary diagnosis*, *working diagnosis*, *differential diagnosis*, *deferred diagnosis*, *undiagnosed illness*, *diagnoses with uncertain or multifactorial etiologies*, *intermittent diagnoses*, *multiple/dual diagnoses*, *self-diagnosis*, or at times *contested diagnosis* need to be part of our vocabulary, thinking, and communications with patients to convey that diagnosis is often imprecise. Anxious patients worried about a condition, for example, cancer, COVID-19 infection, or a diagnosis to which a relative or a friend has recently succumbed, come seeking reassurance and may not welcome an uncertain answer. Thus, we have to work with patients, listen to and respect their concerns, and take their symptoms seriously yet modestly acknowledge our limitations. We need to tailor this approach to patients’ differing levels of health literacy, trust in our clinical advice, and experiences with the health system.

■ DON’T MISS DIAGNOSES AND RED FLAGS

Uncertainty should not be a license for complacency. Particularly for diseases that (1) progress rapidly, (2) require specific treatments that depend on making the correct diagnosis, or (3) have public health or contagion implications, clinicians need to be poised, and systems designed, to consider and, where appropriate, pursue critical “don’t miss” diagnoses. While clinicians are generally aware of more common “don’t miss” diagnoses (e.g., acute myocardial infarction, sepsis), **Table 9-4** illustrates examples of less common diagnoses that warrant similar consideration. Throughout this textbook, readers should orient themselves to recognize such critical diagnoses and think about presentations and syndromes where they may be lurking.

An important related concept is so-called “red flags” or “alarm symptoms.” This construct has its origins in guidelines for back pain but has increasingly been applied to many other problems, such as headache, red eye, swollen joint, or even abdominal pain and chest pain. Examples of widely cited red flags for back pain that should trigger consideration of more serious etiologies include fever, weight loss, history of malignancy or intravenous drug use, or neurologic signs and symptoms. In theory, many presenting syndromes could benefit from identification of such clues to more serious diagnoses. Evidence-based medicine calls for better data on the sensitivity, specificity, yield, and discriminatory ability of various clinical “red flag” clues; yet, few have been rigorously evaluated. Nonetheless, clinicians find them useful as simple ways to reassure themselves and their patients that a common symptom such as back pain or headache is, or is not, likely an indicator of more urgent or serious pathology.

Interwoven with the challenges of not missing critical diagnoses is the problem of overtesting and overdiagnosis—performing unnecessary and even potentially harmful tests whose benefit does not justify the risks or costs or that may lead to diagnoses that would have never

caused any symptoms or problems. Thoughtful diagnosticians need to weigh carefully this “other side of the coin” of missed diagnosis to avoid such harms and expenses.

■ DIAGNOSTIC PITFALLS

One of the important ways of learning in medicine is learning from the missteps of those who have walked the path ahead of us. By learning about commonly missed diagnoses and the ways accurate, timely diagnosis went astray, we can avoid making similar mistakes. Anticipating the potential for similar types of errors can both create situational awareness of traps to avoid and contribute to learning from our own personal and collective patterns of mistakes. Several studies have examined common or recurring pitfalls in diagnosis. An example of a common disease-specific diagnostic pitfall in breast cancer diagnosis is ordering a mammogram for a woman with a palpable breast lump and, when the mammogram returns as normal, reassuring her that cancer has been “ruled out” by the negative test. Any mass or lesion palpable

TABLE 9-4 Examples of “Don’t Miss” Diagnoses

INFECTIONS/ INFLAMMATION	CARDIAC/ISCHEMIC/ BLEEDING	METABOLIC/ HEMATOLOGIC/ ENVIRONMENTAL
Spinal epidural abscess	Aortic dissection Leaking/ruptured abdominal aortic aneurysm	Diabetes ketoacidosis Hyperosmolar hyperglycemia
Necrotizing fasciitis	Pericardial tamponade	Myxedema/thyrototoxicosis
Meningitis	Wolff-Parkinson-White Prolonged QT	Addison’s disease
Endocarditis	Pulmonary embolism	B ₁₂ deficiency anemia
Peritonsillar abscess	Tension pneumothorax	von Willebrand’s disease
Tuberculosis-active pulmonary, other	Acute mesenteric ischemia Sigmoid volvulus	Hemochromatosis
COVID-19 infection	Esophageal, bowel perforation	Celiac sprue
Guillain-Barré syndrome	Cerebellar hemorrhage	Carbon monoxide poisoning
Ebola infection	Spinal cord compression	Food poisoning
Temporal arteritis	Testicular, ovarian torsion	Malignant hyperthermia
Rhabdomyolysis	Ectopic pregnancy	Alcohol, benzodiazepine, barbiturate withdrawal
Angioedema	Retroperitoneal hemorrhage	Tumor lysis syndrome Hypo-/hypercalcemia

TABLE 9-5 Generic Types of Diagnostic Pitfalls

PITFALL	EXAMPLES
Disease A mistaken for disease B Diseases often mistaken/misdiagnosed with each other	<ul style="list-style-type: none"> Aortic dissection misdiagnosed as acute myocardial infarction Bipolar disorder misdiagnosed as depression
Misinterpretation of test result(s) False-positive or false-negative results with failure to recognize test limitations	<ul style="list-style-type: none"> Breast lump dismissed after negative mammogram Negative COVID-19 test early or late in course
Failure to recognize atypical presentation, signs, and symptoms	<ul style="list-style-type: none"> Apathetic hyperthyroidism Sepsis in elderly patient who is afebrile or hypothermic
Failure to assess appropriately the urgency of diagnosis Urgency of the clinical situation was not appreciated and/or delays critical diagnoses	<ul style="list-style-type: none"> Compartment syndrome Pericardial tamponade Tension pneumothorax
Perils of intermittent symptoms or misleading evolution Intermittent symptoms dismissed due to normal findings (exam, lab, electrocardiogram) when initially seen	<ul style="list-style-type: none"> “Lucid interval” in traumatic epidural hematoma Paroxysmal arrhythmias Intermittent hydrocephalus (Bruns’ syndrome)
Confusion arising from response/masking by empiric treatment	<ul style="list-style-type: none"> Empiric treatment with steroids, proton pump inhibitors, antibiotics, pain medication erroneously masking serious diagnosis
Chronic disease or comorbidity presumed to account for new symptoms Especially in medically complex patients	<ul style="list-style-type: none"> Septic joint signs misattributed to chronic rheumatoid arthritis Mental status change due to infection or medication misattributed to underlying dementia
Rare diagnosis: failure to consider or know	<ul style="list-style-type: none"> Many; fortunately, by definition, rare, but still warrant consideration especially if urgent or treatable
Drug or environmental factor not considered/overlooked Underlying etiology causing/contributing to symptoms, or disease progression not sought, uncovered	<ul style="list-style-type: none"> Ventricular arrhythmia related to QT-prolonging drug Achilles tendon rupture related to quinolone drugs
Failure to appreciate risk factors for particular disease	<ul style="list-style-type: none"> Family history of breast, colorectal cancer not solicited and/or weighed in diagnostic evaluation or screening
Failure to appreciate limitations of physical exam Now with ↑ telemedicine, missing physical exam entirely	<ul style="list-style-type: none"> Overweighing absence of tenderness, swelling in deep vein thrombosis Missing pill-rolling tremor during telemedicine visit

on physical examination probably needs more careful assessment proceeding all the way to invasive biopsy, if necessary. Diagnostic pitfalls can be classified into a number of generic scenarios (**Table 9-5**). We now have large databases that have the potential to track “diagnoses outcomes”—i.e., whether a new diagnosis emerges that suggests an initial diagnosis was incorrect or a diagnosis of a patient’s symptoms was suboptimally delayed. This should, in the future, allow us to more rigorously focus on these cases, to identify contributing factors and recurring patterns, and to help point the way for systemwide improvement strategies.

■ DIAGNOSIS SAFETY CULTURE

Just as diagnosing bacterial infections relies on a proper culture medium to grow and identify etiologic organisms, good diagnosis also requires a healthy safety culture that will allow it to grow and flourish. While clinicians may be inclined to view “safety culture” as something too subjective to be important in their quest to make a definitive diagnosis, this view is misguided. Multiple studies have demonstrated adverse consequences resulting from organizational cultures that inhibit openness, learning, and sharing and create a climate where staff

and patients are afraid to speak up when they observe problems or have questions. Most importantly, patients need to be encouraged to question diagnoses and be heard, particularly when they are not responding to treatment as expected or developing symptoms that are either not consistent with the diagnosis or represent possible red flags for other diagnoses or complications.

Studies examining “high-reliability organizations” outside of medicine and “learning health care organizations” have distilled a series of fundamental properties that are correlated with more reliable and safer outcomes. Just as a thermometer or recording of a pulse can suggest how ill a patient is, we now have instruments that can measure safety culture. These safety measurement tools typically are validated staff surveys that assess (1) communication about errors with staff willingness to report mistakes because they do not feel these mistakes are held against them; (2) openness and encouragement to talk about hospital/office problems; (3) existence of a learning culture that seeks to learn from errors and improve based on lessons learned; (4) leadership commitment to safety, prioritizing safety over production speed and the “bottom line” by providing adequate staffing and resources to operate safely; and (5) accountability and transparency for following up safety events and concerns. Each of these generic culture attributes translates into specific implications for diagnostic safety. These include the following:

- Making it “safe” for clinicians to admit and share diagnostic errors
- Proactive identification, ownership, and accountability regarding error-prone diagnostic workflow processes (particularly around test results, referrals, and patient follow-up)
- Leadership making diagnosis improvement a top priority based on recognition that patients and malpractice insurers report that diagnostic errors are the leading patient safety problem
- Mutual trust and respect for challenges that clinicians often face in making diagnoses and caution in applying the lens of hindsight bias in judging what in retrospect might seem like an “obvious” diagnosis that a clinician initially missed

■ HEALTH INFORMATION TECHNOLOGY AND THE FUTURE OF DIAGNOSIS

Clinicians now spend more time interacting with computers than they do interacting with patients. This is especially true for diagnosis and will likely be even more so in the future. Interactions with patients, consultants, and other staff are increasingly mediated through the computer. Key activities, such as collecting patients’ history (past and current), interpreting data to make a diagnosis, conveying diagnostic assessments (to others on the team and, increasingly, to the patient via open notes), and tracking diagnostic trajectories as they evolve over time, are now computer based. With the rise of telemedicine, even elements of the physical examination have been rerouted to electronic encounters.

While many complain the computer has “gotten in the way” of good diagnosis, distracting clinicians from quality time listening to patients and miring doctors in reading and writing notes filled with copied/pasted/templated information of questionable currency and accuracy, medicine needs to harness the computer’s capabilities to improve diagnosis (**Table 9-6**). Although these basic diagnosis-supporting capabilities should be the foundation of the design of health information technology and everyday workflow, electronic medical records have historically been largely designed around other needs, such as ordering medications and billing and malpractice documentation. They need to be radically redesigned to better support diagnostic processes, as well as save, rather than squander, clinicians’ time.

■ DIAGNOSIS OF DIAGNOSIS ERRORS AND SAFETY: PRACTICAL CONCLUSIONS

In practice, there are frequent and meaningful opportunities for improving diagnosis in each of the three NAM-defined areas to make it a) more reliable, b) timely, and c) to improve diagnosis-related communication with patients. Clinicians in training, practicing physicians, nurses, and others should develop the habit of regularly asking

FUNCTION	EXAMPLES
Facilitate collection/gathering of information	<ul style="list-style-type: none"> Quickly access past history from prior care at same and outside institutions Electronic collection of history of present illness, review of systems, and social determinant risks in advance of visits
Enhanced information entry, organization, and display	<ul style="list-style-type: none"> Visually enhanced flowsheets showing trends, relationships to treatment Reorganized notes to facilitate summarization and simplification and prevent items from getting lost
Generating differential diagnosis	<ul style="list-style-type: none"> Automated creation of lists of diagnoses to consider based on patient's symptoms, demographics, risks
Weighing diagnoses likelihoods	<ul style="list-style-type: none"> Tools to assist in calculation of posttest (Bayesian) probabilities
Aids for formulating diagnostic plan, intelligent test ordering	<ul style="list-style-type: none"> Entering a diagnostic consideration (e.g., celiac disease, pheochromocytoma) and computer suggests most appropriate diagnostic test(s) and how to order
Access to diagnostic reference information	<ul style="list-style-type: none"> Info-buttons instantly linking symptom or diagnosis relevant questions to Harrison's, Up-to-Date chapters, references
Ensuring more reliable follow-up	<ul style="list-style-type: none"> Hardwiring "closed loops" to ensure abnormal labs, missed referrals, worrisome symptoms are tracked and followed up
Support screening for early detection	<ul style="list-style-type: none"> Collaborative tools that patients, clinicians, and offices can use to know when due, order and track screening based on individualized demographics, risk factors, prior tests
Collaborative diagnosis; access to specialist	<ul style="list-style-type: none"> Real-time posing/answering of questions Electronic consults; virtual co-management
Facilitating feedback on diagnoses	<ul style="list-style-type: none"> Feeding back new diagnoses (from downstream providers, patients) that emerge suggesting potential misdiagnosis/errors to clinicians, ERs who saw patient previously

Abbreviation: ERs, emergency rooms.

Source: Modified from G Schiff, DW Bates: N Engl J Med 362:1066, 2010, and R El-Karab et al: BMJ Qual Saf Suppl 2:i40, 2013.

themselves three questions about individual patients in their care, and another three questions regarding the systems in which they work. For each patient being assessed, clinicians should ask:

1. What else might this be? (forcing a differential diagnosis to be made)
2. What doesn't fit? (making sure unexplained abnormal findings are not dismissed)
3. What critical diagnoses are important not to miss? (injecting consideration of "don't miss" diagnoses, red flags, and known pitfalls)

and to diagnose safely, each practitioner must recognize that he or she is working within a larger system. Questions to be asking continually, ensuring we are maximizing reliability and timeliness and minimizing potential for errors, include:

1. Do we have reliable "closed loop" systems to provide reliable, ideally automated tracking and following up of patients' symptoms, abnormal laboratory or imaging findings, and critical referrals that we order?
2. What is the culture-of-safety climate in our organization, office, or clinic?
3. How does the electronic (or even paper) medical record as currently implemented help versus impair efficient, timely, accurate, and fail-safe diagnosis, and how can it be improved?

To take these questions to the next stage, an international movement dedicated to studying and improving diagnosis has emerged. These efforts include annual conferences of clinicians, researchers, and patients; the formation of the Society for Improving Diagnosis in Medicine (SIDM); and convening of a broad coalition of organizations,

including the American Board of Internal Medicine (ABIM), the American College of Physicians (ACP), and the Society of Hospital Medicine (SHM), committed to increasing awareness and action. Ultimately, collectively tackling the challenges of improving the quality of diagnosis will transform the way clinicians and patients work together to co-produce better diagnoses.

FURTHER READING

GANDHI TK, SINGH H: Reducing the risk of diagnostic error in the COVID-19 era. J Hosp Med 15:363, 2020.

GRABER ML et al: The impact of electronic health records on diagnosis. Diagnosis (Berl) 4:211, 2017.

NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE. 2015.

Improving Diagnosis in Health Care. <https://doi.org/10.17226/21794>. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, DC.

NEWMAN-TOKER DE et al: Serious misdiagnosis-related harms in malpractice claims: The "big three"—Vascular events, infections, and cancers. Diagnosis (Berl) 6:227, 2019.

SCHIFF GD et al: Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project, in *Advances in Patient Safety: From Research to Implementation. Vol 2: Concepts and Methodology*. Rockville, MD, Agency for Healthcare Research and Quality, 2005.

SCHIFF GD et al: Ten principles for more conservative, care-full diagnosis. Ann Intern Med 169:643, 2018.

10

Racial and Ethnic Disparities in Health Care

Lenny López, Joseph R. Betancourt

Over the course of its history, the United States has experienced dramatic improvements in overall health and life expectancy, largely as a result of initiatives in public health, health promotion, disease prevention, and chronic care management. Our ability to prevent, detect, and treat diseases in their early stages has allowed us to target and reduce rates of morbidity and mortality. Despite interventions that have improved the overall health of the majority of Americans, racial and ethnic minorities (blacks, Hispanics/Latinos, Native Americans/Alaskan Natives, Asian/Pacific Islanders) have benefited less from these advances than whites and have suffered poorer health outcomes from many major diseases, including cardiovascular disease, cancer, and diabetes. These disparities highlight the importance of recognizing and addressing the multiple factors that impact health outcomes, including structural racism, *social determinants of health* (SDOH), access to care, and health care quality. On this last point, research has revealed that minorities may receive less care and lower-quality care than whites, even when confounders such as stage of presentation, comorbidities, and health insurance are controlled. These differences in quality are called *racial and ethnic disparities in health care*. These health care disparities have taken on greater importance with the significant transformation of the U.S. health care system and value-based purchasing. The shift toward creating financial incentives and disincentives to achieve quality goals makes focusing on those who receive lower-quality care more important than ever before. This chapter will provide an overview of racial and ethnic disparities in health and health care, identify root causes, and provide key recommendations to address these disparities at both the clinical and health system levels.

NATURE AND EXTENT OF DISPARITIES

Life expectancy at birth is an important measure of the health of a nation's population. Although the overall life expectancy in the United States has been increasing since 1900, differences due to

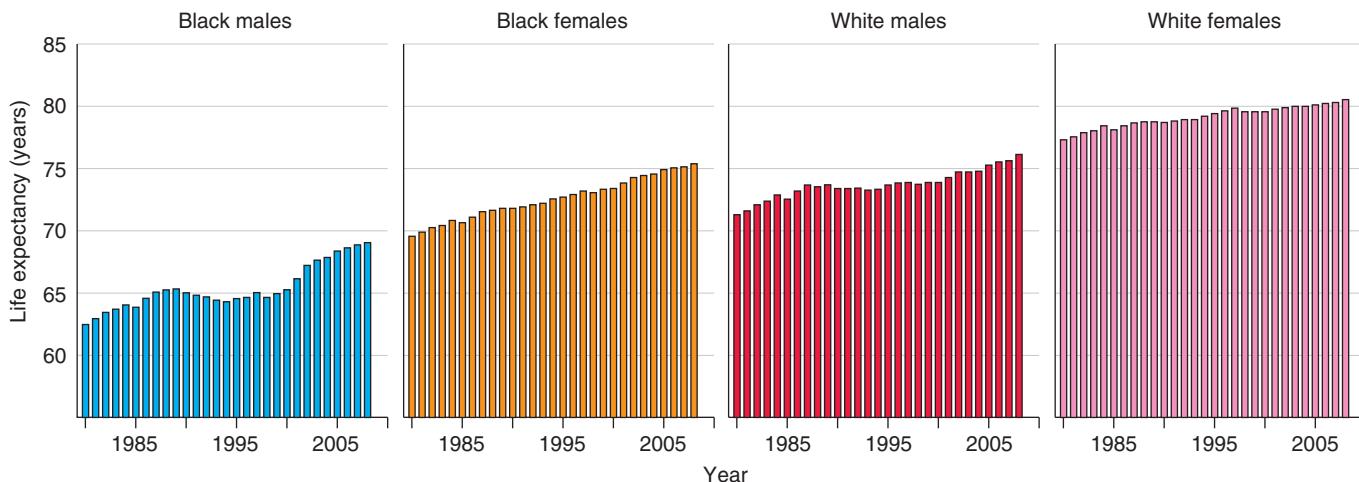


FIGURE 10-1 Life expectancy at birth among black and white males and females in the United States, 1975–2003. (Adapted from S Harper, J Lynch, S Burris, GD Smith: Trends in the black-white life expectancy gap in the United States, 1983–2003. *JAMA* 297:1224, 2007.)

race/ethnicity, education, and socioeconomic status have persisted. For example, at every level of education and income, African Americans have lower life expectancy at age 25 than whites and Hispanics/Latinos. Blacks with a college degree or more education have lower life expectancy than whites and Hispanics who graduated from high school. Blacks have had lower life expectancy compared to whites for as long as data have been collected. From 1975 to 2003, the largest difference in life expectancy between blacks and whites was substantial (6.3 years for males and 4.5 years for females) (Fig. 10-1). The gap in life expectancy between the black and white populations decreased by 2.3 years between 1999 and 2013 from 5.9 to 3.6 years (4.4 years for males and 3.0 years for women) (Fig. 10-2).

The life expectancy gap is augmented by worse health and higher disease burden. Cardiovascular-related diseases remain the leading cause of black-white differences in life expectancy. If all cardiovascular causes and diabetes are considered together, they account for 35% and 52% of the gap for males and females, respectively. Finally, place matters for health. Analysis of data from 2010 to 2015 demonstrate large geographic life expectancy gap variation at the census tract level (Fig. 10-3). Socioeconomic and race/ethnicity factors, behavioral and metabolic risk factors (prevalence of obesity, leisure-time physical inactivity, cigarette smoking, hypertension, diabetes), and health care factors (percentage of the population younger than 65 years who are insured, primary care access and quality, number of physicians per capita) explained 60%, 74%, and 27% of county-level variation in life expectancy, respectively. Combined, these factors explained 74% of this variation. Most of the association between socioeconomic and race/ethnicity factors and life expectancy was mediated through behavioral and metabolic risk factors.

In addition to racial and ethnic disparities in *health*, there are racial and ethnic disparities in the *quality of care* for persons with access to

the health care system. Seminal studies over several decades have consistently documented disparities in health care. For instance, studies have documented disparities in the treatment of pneumonia and congestive heart failure, with blacks receiving less optimal care than whites when hospitalized for these conditions. Moreover, blacks with end-stage renal disease are referred less often to the transplant list than are their white counterparts (Fig. 10-4). Disparities have been found, for example, in the use of cardiac diagnostic and therapeutic procedures (with blacks being referred less often than whites for cardiac catheterization and bypass grafting), prescription of analgesia for pain control (with blacks and Hispanics/Latinos receiving less pain medication than whites for long-bone fractures and cancer), and surgical treatment of lung cancer (with blacks receiving less curative surgery than whites for non-small-cell lung cancer). Again, many of these disparities have occurred even when variations in factors such as insurance status, income, age, comorbid conditions, and symptom expression are taken into account. Finally, disparities in the quality of care provided at the sites where minorities tend to receive care have been shown to be an important additional contributor to overall disparities.

The 2019 National Healthcare Quality and Disparities Report, released by the Agency for Healthcare Research and Quality, tracks about 250 health care process, outcome, and access measures, across many diseases and settings. This annual report is particularly important because most studies of disparities have not been longitudinally repeated with the same methodology to document trends and changes in disparities over time. This report found that some disparities were getting smaller from 2000 through 2016–2018, but disparities persisted and some even worsened, especially for poor and uninsured populations. For about 40% of quality measures, blacks (82 of 202 measures) and American Indians and Alaska Natives (47 of 116 measures) received worse care than whites. For more than one-third of quality measures, Hispanics (61 of 177 measures) received worse care than whites. Asians and Native Hawaiians/Pacific Islanders received worse care than whites for about 30% of quality measures, but Asians also received better care for about 30% of quality measures (Fig. 10-5). Of note, for those quality measures that demonstrated disparities at baseline, >90% of these measures showed no improvement since 2000 (Fig. 10-6).

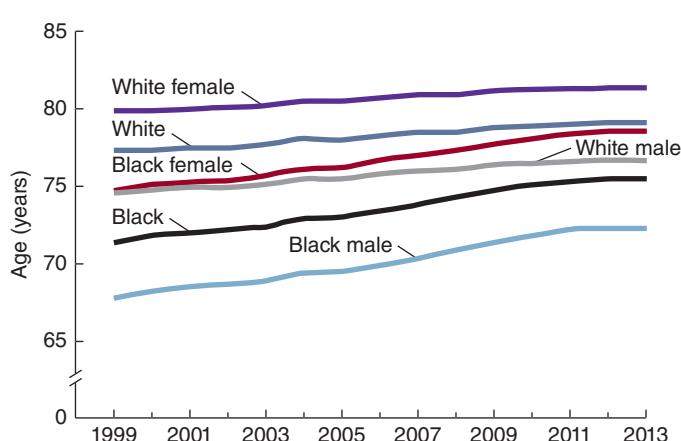


FIGURE 10-2 Life expectancy, by race and sex: United States, 1999–2013. (From KD Kochanek et al: NCHS Data Brief 218:1, 2015.)

■ ROOT CAUSES OF DISPARITIES

Race, Racism, and Health Race and racism are core elements of any explanatory model on racial and ethnic disparities in health and health care. Our nation's history of slavery, segregation, separate but "equal" health care, and medical experimentation, among a myriad of other ways in which racism has manifested in the United States, has played a key role in the existence and persistence of these disparities. It is now well accepted that race is a social category without biologic foundation and a product of historical racism. Nevertheless, it is clear that racism has a biologic impact as a form of psychosocial stress. It is now well established

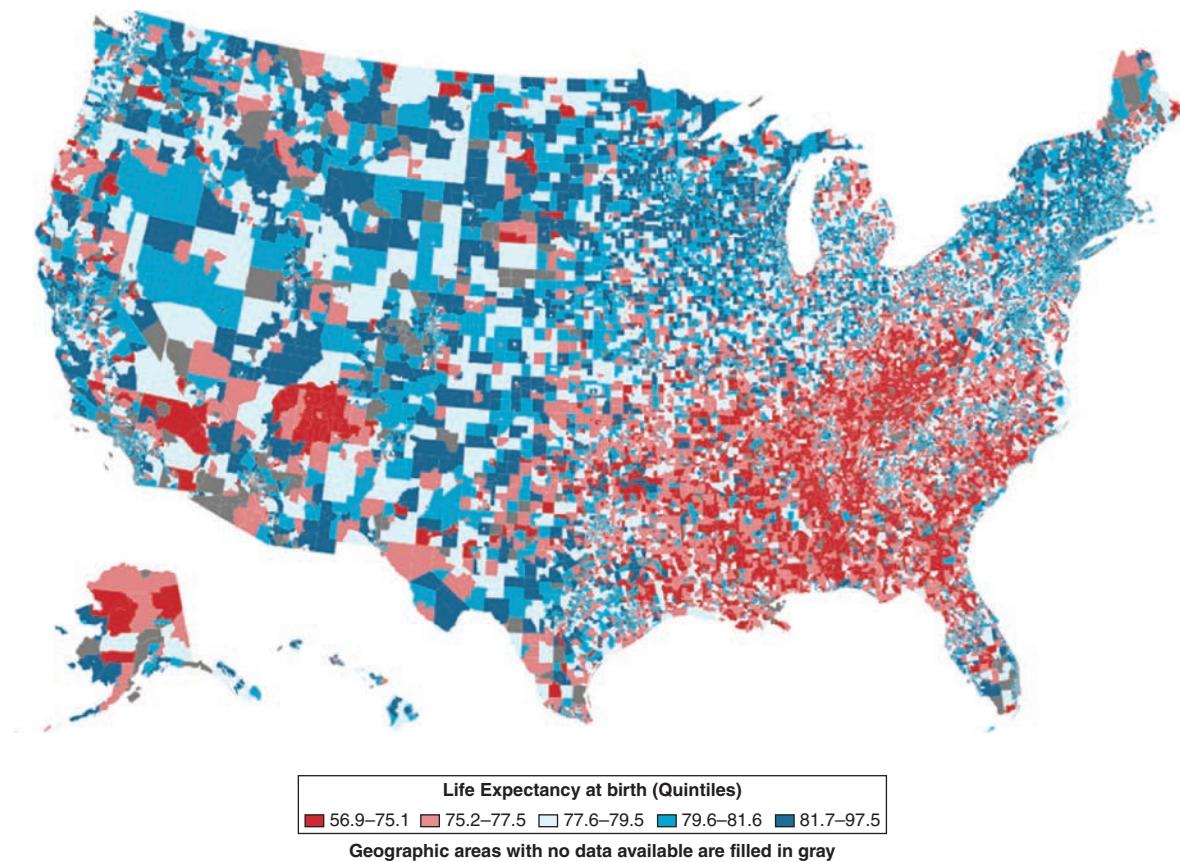


FIGURE 10-3 Life expectancy at birth for U.S. census tracts, 2010-2015. (A New View of Life Expectancy, Surveillance and Data - Blogs and Stories, Centers for Disease Control and Prevention. Retrieved from <https://www.cdc.gov/surveillance/blogs-stories/life-expectancy.html>.)

that psychosocial stress negatively impacts health through psychophysiological reactivity causing hyperstimulation of the sympathetic-adrenal-medullary system and the hypothalamic-pituitary-adrenal axis, leading to vascular inflammation, endothelial dysfunction, and neurohormonal dysregulation causing an acceleration of cardiovascular disease. Behavioral changes occurring as adaptations or coping responses to stressors such as increased smoking, decreased exercise and sleep, and

poorer adherence to medical regimens provide an additional important pathway through which stressors influence disease risk. This accelerated disease risk, aging, and premature death has been termed the *weathering effect*.

While most empiric research focuses on interpersonal racial/ethnic discrimination, structural racism (sometimes called institutional racism) provides a more holistic framework. Structural racism refers to the totality of ways that a society fosters, sustains, and reinforces discrimination through sociopolitical, legal, economic, and health structures that determine differential access to risks, opportunities, and resources that drive health and health care disparities. Structural racism explains how racism's structure and ideology can persist in governmental and institutional policies in the absence of individual actors who are explicitly racially prejudiced. For example, the history of residential segregation has had lasting negative effects generationaly on equal access for racial/ethnic minorities to employment, banking, earnings, high-quality education, and health care. Policies that do not address root structural causes will not address health and health care inequities.

With the promise of individualizing clinical decisions, the use of race in clinical and risk assessment algorithms has long been a part of modern medicine. The evidence is now clear that race is not a reliable proxy for genetic difference and that race adjustment has the potential to create inadvertent disparities in health care. One clinical example is from nephrology. Blacks have higher rates of end-stage kidney disease and death due to kidney failure than the overall population. The most widely used cohort-derived equation to estimate glomerular filtration rate (GFR), the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation, has the limitation that it produces 80–90% estimated GFR (eGFR) values that are within $\pm 30\%$ of a patient's measured GFR. In addition, this equation uses a black race-related factor, which increases eGFR for any given serum creatinine by 15.9% compared to a nonblack patient with the same age, sex, and serum creatinine. The increase in eGFR is likely to disadvantage blacks for early referral to a nephrologist, early treatment of advanced chronic kidney disease, and

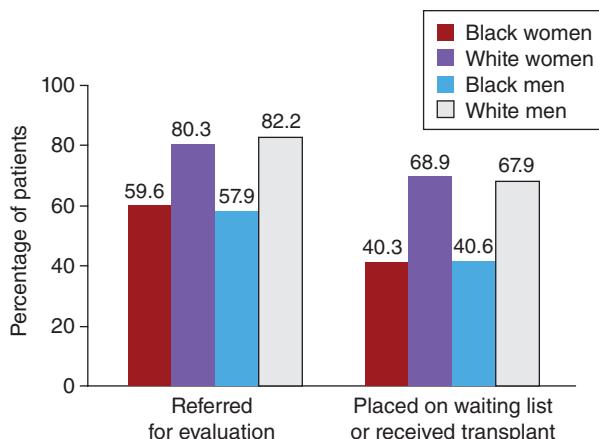


FIGURE 10-4 Referral for evaluation at a transplantation center or placement on a waiting list/receipt of a renal transplant within 18 months after the start of dialysis among patients who wanted a transplant, according to race and sex. The reference population consisted of 239 black women, 280 white women, 271 black men, and 271 white men. Racial differences were statistically significant among both the women and the men ($p < .0001$ for each comparison). (From JZ Ayanian, PD Cleary, JS Weissman, AM Epstein: The effect of patients' preferences on racial differences in access to renal transplantation. *N Engl J Med* 341:1661, 1999. Copyright © 1999 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

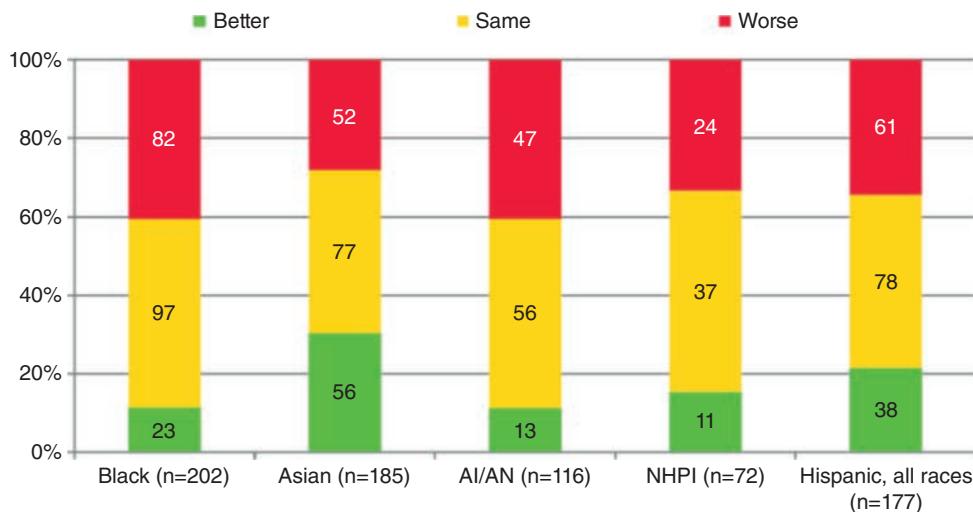


FIGURE 10-5 Number and percentage of quality measures for which members of selected groups experienced better, same, or worse quality of care compared with reference group (white) for the most recent data year, 2014, 2016, 2017, or 2018. AI/AN, American Indian or Alaska Native; NHPI, Native Hawaiian/Pacific Islander (From 2019 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality; December 2020. AHRQ Pub. No. 20(21)-0045-EF.)

kidney transplantation. It is also not clear how to apply the race factor when the patient's race is unknown and/or ambiguous, as in those who are multiracial. This disparity-inducing scenario could be avoided through the use of cystatin C-based eGFR estimation, which has been demonstrated to be more accurate than the CKD-EPI equation and for which race is not required in estimation.

The application of artificial intelligence (AI) analytics to large amounts of clinical electronic data—big data—holds the promise to better understand health care costs, utilization, resource allocation, and population health monitoring. Machine learning models can identify the statistical patterns in large amounts of historically collected data. These data naturally contain the patterning of preexisting health care disparities created by socially and historically structured inequities. This biased patterning can lead to incorrect predictions, withholding of resources, and worse outcomes for vulnerable populations. Recently, analysis of a commercial, national, proprietary prediction algorithm, affecting millions of patients, exhibited racial bias. Historical cost data were used to predict clinical risk and allocate additional clinical services for high-cost patients. Algorithmic bias arose because black patients historically have less access to health care and thus less money is spent on their care compared to white patients. Thus, blacks, who tended to be sicker than white patients, received lower clinical risk scores and thus were less likely to receive additional clinical services. The observed allocation bias was remedied using direct measures of illness and illness severity. Thus, machine learning algorithms are not inherently free of bias and should be assessed for accuracy and fairness.

In summary, there are many ways in which racism has contributed and does and will continue to contribute to racial and ethnic disparities in health and health care.

SOCIAL DETERMINANTS OF HEALTH

Minority Americans have poorer health outcomes than whites from preventable and treatable conditions such as cardiovascular disease, diabetes, asthma, cancer, and HIV/AIDS. Multiple factors contribute to these racial and ethnic disparities in health. The landmark

National Academy of Medicine (formerly, the Institute of Medicine [IOM]) report, *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, published in 2002, summarized the scientific evidence on health disparities and provided an important framework for conceptualizing and defining racial/ethnic disparities. Since the *Unequal Treatment* report, there has been a growing empirical evidence base on how racism and the SDOH, often working in synergy, create and sustain disparities. Mechanistically, the biopsychosocial model brings together the social and physical characteristics of the environment with individual physical and psychological attributes. These environmental and individual characteristics, in turn, influence health behaviors and stress-related physiologic pathways that directly impact health. The National Institute on Minority Health and Health Disparities SDOH model builds on prior models and adds

the time element across the life course of the individual in recognition of the long-lasting health effects of socioeconomic exposures (Fig. 10-7). The resulting matrix has the domains of influence of health (biological, behavioral, physical and built environment, sociocultural environment, health care system) along the y-axis and the levels of influence on health (individual, interpersonal, community, societal) along the x-axis. Cells are not mutually exclusive, and examples of factors within each cell are illustrative and not comprehensive. This framework emphasizes the complex multidomain etiologies of disparities across the factors in the conceptual matrix thus highlighting the limitation of individual-level focused research and policy.

In addition to race and racism, *Unequal Treatment* identified a set of root causes that included health system, provider-level, and patient-level factors.

Health System Factors • HEALTH SYSTEM COMPLEXITY Even among persons who are insured and educated and who have a high degree of health literacy, navigating the U.S. health care system can be complicated and confusing. Some individuals may be at higher risk for receiving substandard care because of their difficulty navigating the system's complexities. These individuals may include those from cultures

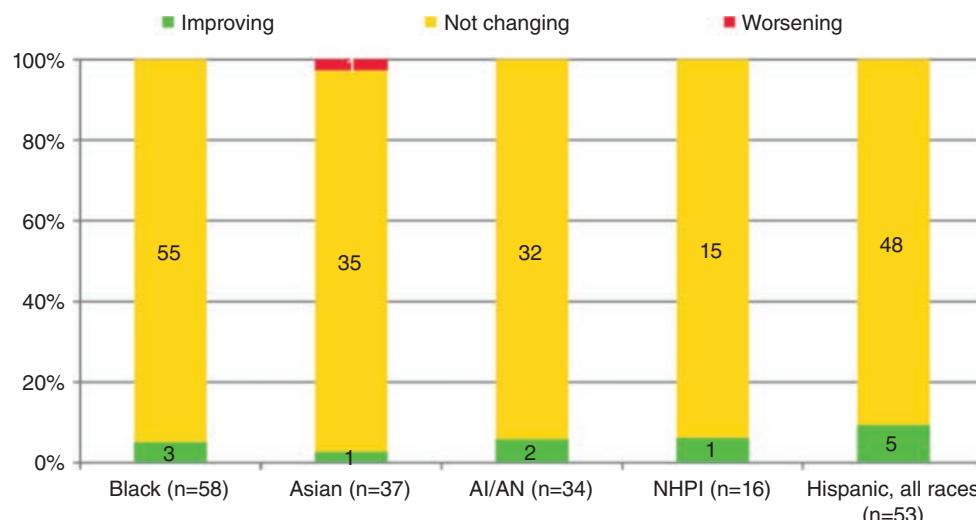


FIGURE 10-6 Number and percentage of quality measures with disparity at baseline for which disparities related to race and ethnicity were improving, not changing, or worsening over time, 2000 through 2014, 2015, 2016, 2017, or 2018. AI/AN, American Indian or Alaska Native; NHPI, Native Hawaiian/Pacific Islander. (From 2019 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality; December 2020. AHRQ Pub. No. 20(21)-0045-EF.)

		Levels of Influence*			
		Individual	Interpersonal	Community	Societal
Domains of Influence (Over the Lifecourse)	Biological	Biological Vulnerability and Mechanisms	Caregiver–Child Interaction Family Microbiome	Community Illness Exposure Herd Immunity	Sanitation Immunization Pathogen Exposure
	Behavioral	Health Behaviors Coping Strategies	Family Functioning School/Work Functioning	Community Functioning	Policies and Laws
	Physical/Built Environment	Personal Environment	Household Environment School/Work Environment	Community Environment Community Resources	Societal Structure
	Sociocultural Environment	Sociodemographics Limited English Cultural Identity Response to Discrimination	Social Networks Family/Peer Norms Interpersonal Discrimination	Community Norms Local Structural Discrimination	Social Norms Societal Structural Discrimination
	Health Care System	Insurance Coverage Health Literacy Treatment Preferences	Patient–Clinician Relationship Medical Decision-Making	Availability of Services Safety Net Services	Quality of Care Health Care Policies
	Health Outcomes	 Individual Health	 Family/ Organizational Health	 Community Health	 Population Health

FIGURE 10-7 National Institute on Minority Health and Health Disparities social determinants research framework.*Health disparity populations: race/ethnicity, low socioeconomic status, rural, sexual and gender minority. Other fundamental characteristics: sex and gender, disability, geographic region. (From National Institute on Minority Health and Health Disparities. NIMHD Research Framework. 2017. Retrieved from <https://www.nimhd.nih.gov/about/overview/research-framework.html>.)

unfamiliar with the Western model of health care delivery, those with limited English proficiency, those with low health literacy, and those who are mistrustful of the health care system. These individuals may have difficulty knowing how and where to go for a referral to a specialist; how to prepare for a procedure such as a colonoscopy; or how to follow up on an abnormal test result such as a mammogram. Since people of color in the United States tend to be overrepresented among the groups listed above, the inherent complexity of navigating the health care system has been seen as a root cause for racial/ethnic disparities in health care.

OTHER HEALTH SYSTEM FACTORS Racial/ethnic disparities are due not only to differences in care provided within hospitals but also to where and from whom minorities receive their care; i.e., certain specific providers, geographic regions, or hospitals are lower-performing on certain aspects of quality. For example, one study showed that 25% of hospitals cared for 90% of black Medicare patients in the United States and that these hospitals tended to have lower performance scores on certain quality measures than other hospitals. That said, health systems generally are not well prepared to measure, report, and intervene to reduce disparities in care. Few hospitals or health plans stratify their quality data by race/ethnicity or language to measure disparities, and even fewer use data of this type to develop disparity-targeted interventions. Similarly, despite regulations concerning the need for professional interpreters, research demonstrates that many health care organizations and providers fail to routinely provide this service for patients with limited English proficiency. Despite the link between limited English proficiency and health care quality and safety, few providers or institutions monitor performance for patients in these areas.

Provider-Level Factors • PROVIDER-PATIENT COMMUNICATION

Significant evidence highlights the impact of sociocultural factors, race, ethnicity, and limited English proficiency on health and clinical care. Health care professionals frequently care for diverse populations with varied perspectives, values, beliefs, and behaviors regarding health and well-being. The differences include variations in the recognition of symptoms, thresholds for seeking care, comprehension of management strategies, expectations of care (including preferences for or against diagnostic and therapeutic procedures), and adherence to preventive

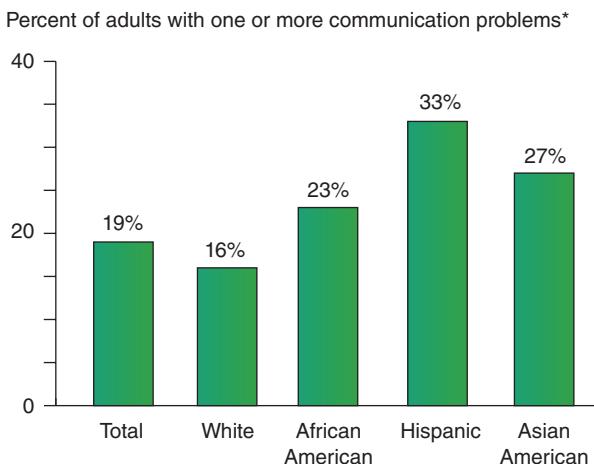
measures and medications. In addition, sociocultural differences between patient and provider influence communication and clinical decision-making and are especially pertinent: evidence clearly links provider–patient communication to improved patient satisfaction, regimen adherence, and better health outcomes (Fig. 10-8). Thus, when sociocultural differences between patient and provider are not appreciated, explored, understood, or communicated effectively during the medical encounter, patient dissatisfaction, poor adherence, poorer health outcomes, and racial/ethnic disparities in care may result.

A survey of 6722 Americans ≥18 years of age is particularly relevant to this important link between provider–patient communication and health outcomes. Whites, African Americans, Hispanics/Latinos, and Asian Americans who had made a medical visit in the past 2 years were asked whether they had trouble understanding their doctors; whether they felt the doctors did not listen; and whether they had medical questions they were afraid to ask. The survey found that 19% of all patients experienced one or more of these problems, yet whites experienced them 16% of the time as opposed to 23% of the time for African Americans, 33% for Hispanics/Latinos, and 27% for Asian Americans (Fig. 10-9).

How do we link communication to outcomes?



FIGURE 10-8 The link between effective communication and patient satisfaction, adherence, and health outcomes. (Institute of Medicine. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. <https://doi.org/10.17226/12875>. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)



Base: Adults with health care visit in past two years

*Problems include understanding doctor, feeling doctor listened, had questions but did not ask.

FIGURE 10-9 Communication difficulties with physicians, by race/ethnicity. The reference population consisted of 6722 Americans ≥ 18 years of age who had made a medical visit in the previous 2 years and were asked whether they had had trouble understanding their doctors, whether they felt that the doctors had not listened, and whether they had had medical questions they were afraid to ask. (Reproduced with permission from the Commonwealth Fund Health Care Quality Survey, 2001.)

In addition, in the setting of even a minimal language barrier, provider–patient communication without an interpreter is recognized as a major challenge to effective health care delivery. These communication barriers for patients with limited English proficiency lead to frequent misunderstanding of diagnosis, treatment, and follow-up plans; inappropriate use of medications; lack of informed consent for surgical procedures; high rates of adverse events with more serious clinical consequences; and a lower-quality health care experience than is provided to patients who speak fluent English. Physicians who have access to trained interpreters report a significantly higher quality of patient–physician communication than physicians who use other methods. Communication issues related to discordant language disproportionately affect minorities and likely contribute to racial/ethnic disparities in health care.

CLINICAL DECISION-MAKING Theory and research suggest that variations in clinical decision-making may contribute to racial and ethnic disparities in health care. Two factors are central to this process: clinical uncertainty and stereotyping.

First, a doctor's decision-making process is nested in *clinical uncertainty*. Doctors depend on inferences about severity based on what they understand about illness and the information obtained from the patient. A doctor caring for a patient whose symptoms he or she has difficulty understanding and whose “signals”—the set of clues and indications that physicians rely on to make clinical decisions—are hard to read may make a decision different from the one that would be made for another patient who presents with exactly the same clinical condition. Given that the expression of symptoms may differ among cultural and racial groups, doctors—the overwhelming majority of whom are white—may understand symptoms best when expressed by patients of their own racial/ethnic groups. The consequence is that white patients may be treated differently from minority patients. Differences in clinical decisions can arise from this mechanism even when the doctor has the same regard for each patient (i.e., is not prejudiced).

Second, the literature on social cognitive theory highlights how natural tendencies to stereotype may influence clinical decision-making. Stereotyping can be defined as the way in which people use social categories (e.g., race, gender, age) in acquiring, processing, and recalling information about others. Faced with enormous information loads and the need to make many decisions, people often subconsciously simplify the decision-making process and lessen cognitive effort by using “categories” or “stereotypes” that bundle information into groups or

types that can be processed more quickly. Although functional, stereotyping can be systematically biased, as people are automatically classified into social categories based on dimensions such as *race*, *gender*, and *age*. Many people may not be aware of their attitudes, may not consciously endorse specific stereotypes, and paradoxically may consider themselves egalitarian and not prejudiced.

Stereotypes may be strongly influenced by the messages presented consciously and unconsciously in society. For instance, if the media and our social/professional contacts tend to present images of minorities as being less educated, more violent, and nonadherent to health care recommendations, these impressions may generate stereotypes that unnaturally and unjustly impact clinical decision-making. As signs of racism, classism, gender bias, and ageism are experienced (consciously or unconsciously) in our society, stereotypes may be created that impact the way doctors manage patients from these groups. On the basis of training or practice location, doctors may develop certain perceptions about race/ethnicity, culture, and class that may evolve into stereotypes. For example, many medical students and residents are trained—and minorities cared for—in academic health centers or public hospitals located in socioeconomically disadvantaged areas. As a result, doctors may begin to equate certain races and ethnicities with specific health beliefs and behaviors (e.g., “these patients” engage in risky behaviors, “those patients” tend to be noncompliant) that are more associated with the social environment (e.g., poverty) than with a patient's racial/ethnic background or cultural traditions. This “conditioning” phenomenon may also be operative if doctors are faced with certain racial/ethnic patient groups who frequently do not choose aggressive forms of diagnostic or therapeutic intervention. The result over time may be that doctors begin to believe that “these patients” do not like invasive procedures; thus, they may not offer these procedures as options. A wide range of studies have documented the potential for provider biases to contribute to racial/ethnic disparities in health care. For example, one study measured physicians' unconscious (or implicit) biases and showed that these were related to differences in decisions to provide thrombolysis for a hypothetical black or white patient with a myocardial infarction.

It is important to differentiate stereotyping from prejudice and discrimination. *Prejudice* is a conscious prejudgetment of individuals that may lead to disparate treatment, and *discrimination* is conscious and intentional disparate treatment. All individuals *stereotype* subconsciously, yet, if left unquestioned, these subconscious assumptions may lead to lower-quality care for certain groups because of differences in clinical decision-making or differences in communication and patient-centeredness. For example, one study tested physicians' unconscious racial/ethnic biases and showed that patients perceived more biased physicians as being less patient-centered in their communication. What is particularly salient is that stereotypes tend to be activated most in environments where the individual is stressed, multitasking, and under time pressure—the hallmarks of the clinical encounter. In fact, in a survey of close to 16,000 physicians, 42% admitted that bias—including by race and ethnicity—impacted their clinical decision-making. Interestingly, emergency medicine physicians, who work in environments of stress, time pressure, risk, and where they are multitasking, topped the list by discipline at 62%.

Patient-Level Factors Lack of trust has become a major concern for many health care institutions today. For example, an IOM report, *To Err Is Human: Building a Safer Health System*, documented alarming rates of medical errors that made patients feel vulnerable and less trusting of the U.S. health care system. The increased media and academic attention to problems related to quality of care (and of disparities themselves) has clearly diminished trust in doctors and nurses.

Trust is a crucial element in the therapeutic alliance between patient and health care provider. It facilitates open communication and is directly correlated with adherence to the physician's recommendations and the patient's satisfaction. In other words, patients who mistrust their health care providers are less satisfied with the care they receive, and mistrust of the health care system greatly affects patients' use of services. Mistrust can also result in inconsistent care, “doctor-shopping,”

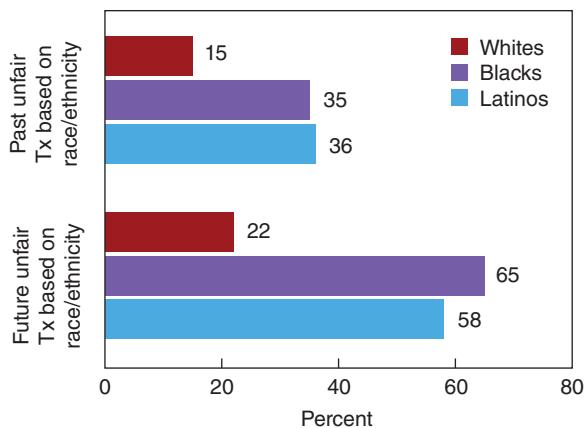


FIGURE 10-10 Patient perspectives regarding unfair treatment (Tx) based on race/ethnicity. The reference population consisted of 3884 individuals surveyed about how fairly they had been treated in the health care system in the past and how fairly they felt they would be treated in the future on the basis of their race/ethnicity. (From *Race, Ethnicity & Medical Care: A Survey of Public Perceptions and Experiences*. Kaiser Family Foundation, 2005.)

self-medication, and an increased demand by patients for referrals and diagnostic tests.

On the basis of historic factors such as discrimination, segregation, and medical experimentation, blacks may be especially mistrustful of providers. The exploitation of blacks by the U.S. Public Health Service during the Tuskegee syphilis study from 1932 to 1972 left a legacy of mistrust that persists even today among this population. Other populations, including Native Americans/Alaskan Natives, Hispanics/Latinos, and Asian Americans, also harbor significant mistrust of the health care system. A national survey conducted by the Kaiser Family Foundation found that there is significant mistrust for the health care system among minority populations. Of the 3884 individuals surveyed, 36% of Hispanics and 35% of blacks (compared to 15% of whites) felt they were treated unfairly in the health care system in the past based on their race and ethnicity. Perhaps even more alarming—65% of blacks and 58% of Hispanics (compared to 22% of whites) were afraid of being treated unfairly in the future based on their race/ethnicity (Fig. 10-10).

This mistrust may contribute to wariness in accepting or following recommendations, undergoing invasive procedures, or participating in clinical research, and these choices, in turn, may lead to misunderstanding and the perpetuation of stereotypes among health professionals.

■ KEY RECOMMENDATIONS TO ADDRESS RACIAL/ETHNIC DISPARITIES IN HEALTH CARE

Unequal Treatment provides recommendations to address the root causes of racial/ethnic disparities organized as *health system interventions*, *provider interventions*, *patient interventions*, and *general recommendations*.

Health System Interventions • COLLECTING, REPORTING, AND TRACKING OF DATA ON HEALTH CARE ACCESS AND USE, BY PATIENTS' RACE/ETHNICITY *Unequal Treatment* found that the appropriate systems to track and monitor racial and ethnic disparities in health care are lacking and that less is known about the disparities affecting minority groups other than African Americans (Hispanics, Asian Americans, Pacific Islanders, Native Americans, and Alaskan Natives). For instance, only in the mid-1980s did the Medicare database begin to collect data on patient groups outside the standard categories of “white,” “black,” and “other.” Federal, private, and state-supported data-collection efforts are scattered and unsystematic, and many health care systems and hospitals still do not collect data on the race, ethnicity, or primary language of enrollees or patients. A survey by the Institute for Diversity in Health Management and the Health Research and Educational Trust in 2015 found that 98% of 1083 U.S. hospitals collected information on race, 95% collected data on ethnicity, and 94% collected data on primary language. However, only 45% collected

data on race, 40% collected data on ethnicity, and 38% collected data on primary language to benchmark gaps in care. A survey by America's Health Insurance Plans Foundation in 2008 and 2010 showed that the proportion of enrollees in plans that collected race/ethnicity data of some type increased from 75 to 79%; however, the total percentage of plan enrollees whose race/ethnicity and language are recorded is still much lower than these figures.

COLLECTING, REPORTING, AND TRACKING OF SDOH DATA In 2014, the IOM Committee on Recommended Social and Behavioral Domains and Measures for Electronic Health Records recommended the routine collection, in the electronic health record, of a parsimonious panel of clinically significant SDOH measures that may be obtained by self-report in advance of or during the health care encounter and, when used together, provide a psychosocial vital sign. The IOM-recommended questionnaire includes 25 items addressing the following domains: race and ethnicity, education, financial resource strain, stress, depression, physical activity, tobacco use, alcohol use, social connection or isolation, intimate partner violence, residential address, and geocoded census tract median income. Implementation studies have demonstrated that collection of these data takes about 5 minutes, and both patients and providers saw this data collection as appropriate and important. Given that data access and monitoring is an essential component to disparities elimination, we highlight several important sources of up-to-date racial/ethnic disparities monitoring initiatives that are available to the general public and are updated regularly. We highlight only three examples of national data sources.

- Since 2003, the Agency for Healthcare Research and Quality has led the yearly compilation of *The National Healthcare Quality and Disparities Report*, which reports trends for measures related to access to health care, affordable care, care coordination, healthy living, patient safety, and the quality of care across acute and chronic disease management by race/ethnicity, income, and other SDOH (<https://www.ahrq.gov/research/findings/nhqrdr/index.html>).
- Since 2011, the Geospatial Research, Analysis, and Services Program (GRASP) created and maintains the Centers for Disease Control and Prevention Social Vulnerability Index. This database maps, for all U.S. Census tracts, 15 social factors (grouped in four SDOH categories: socioeconomic status, housing composition and disability, minority status and language, and housing and transportation) and is updated every 2 years (<https://www.atsdr.cdc.gov/placeandhealth/svi/index.html>).
- Launched in 2018, the Health Opportunity and Equity (HOPE) Initiative benchmarks and tracks 27 indicators by race, ethnicity, and socioeconomic status. The indicators measure social and economic factors, community and safety, physical environment, access to health care, and health outcomes for the United States (<https://www.nationalcollaborative.org/our-programs/hope-initiative-project/>).

INCREASE INSURANCE COVERAGE AND ACCESS Lack of access to high-quality health care is an important driver of racial/ethnic disparities. Signed into law in 2010, the Affordable Care Act (ACA) fundamentally transformed health insurance by decreasing the uninsured population from 16.3% in 2010 (~49.9 million) to 8.8% in 2016 (~28.1 million). This represents the largest expansion of health insurance since the creation of Medicare and Medicaid in 1965. Prior to the ACA, non-Hispanic blacks were 70% and Hispanics nearly three times more likely to be uninsured than non-Hispanic whites. Of note, Medicaid expansion accounted for an estimated 60% of the ACA's effect through a combination of expanded eligibility and increased enrollment of previously eligible but unenrolled people. This is important given the higher number of racial/ethnic minorities who obtain insurance through Medicaid. Many studies have demonstrated that increased insurance coverage has also translated to greater improvement for blacks and Hispanics in access to care, more access to a usual source of care, and improved health outcomes.

ENCOURAGEMENT OF THE USE OF EVIDENCE-BASED GUIDELINES AND QUALITY IMPROVEMENT *Unequal Treatment* highlights the subjectivity of clinical decision-making as a potential cause of racial and

ethnic disparities in health care by describing how clinicians—despite the existence of well-delineated practice guidelines—may offer (consciously or unconsciously) different diagnostic and therapeutic options to different patients on the basis of their race or ethnicity. Therefore, the widespread adoption and implementation of evidence-based guidelines is a key recommendation in eliminating disparities. For instance, evidence-based guidelines are now available for the management of diabetes, HIV/AIDS, cardiovascular diseases, cancer screening and management, and asthma—all areas where significant disparities exist. As part of ongoing quality-improvement efforts, particular attention should be paid to the implementation of evidence-based guidelines for all patients, regardless of their race and ethnicity.

SUPPORT FOR THE USE OF LANGUAGE INTERPRETATION SERVICES IN THE CLINICAL SETTING As described previously, a lack of efficient and effective interpreter services in a health care system can lead to patient dissatisfaction, to poor comprehension and adherence, and thus to ineffective/lower-quality care for patients with limited English proficiency. *Unequal Treatment's* recommendation to support the use of interpretation services has clear implications for delivery of quality health care by improving doctors' ability to communicate effectively with these patients.

INCREASES IN THE PROPORTION OF UNDERREPRESENTED MINORITIES IN THE HEALTH CARE WORKFORCE Data for 2018 from the Association of American Medical Colleges indicate that of active physicians, 56.2% identified as white, 5.8% identified as Hispanic, 5.0% identified as black or African American, and 0.3% identified as Native American or Alaskan Natives. Furthermore, U.S. national data show that only 3.6% of full-time faculty are black or African American, and 5.5% are Hispanic, Latino, or of Spanish origin (alone or in combination with another race/ethnicity), compared to 63.9% who identified as white. Longitudinal data demonstrate that minority faculty are more likely to be at or below the rank of assistant professor, while whites composed the highest proportion of full professors. Similarly, several studies have found that both Hispanic and black faculty were promoted at lower rates than their white counterparts. Despite representing ~30% of the U.S. population (a number projected to almost double by 2050), minority students are still underrepresented in medical schools. In 2018, matriculates to U.S. medical schools were 6.2% Latino, 7.1% African American, 0.1% Native Hawaiian or Other Pacific Islander, and 0.2% Native American or Alaskan Native. These percentages have decreased or remained nearly the same since 2007. It will be difficult to develop a diverse physician workforce that can meet the needs of an increasingly diverse population without dramatic changes in the racial and ethnic composition of medical student bodies. Long-term investment in pipeline programs and the nearly universal adoption of holistic admissions (a process by which schools consider each applicant individually to determine how they might contribute to the learning environment and the workforce instead of relying just on test scores and grades) have produced modest results. Institutional change in medical schools, focused on creating nurturing, inclusive, and equity-focused environments that dismantle the structural racism that has created the opportunity gap faced by many minority students, is needed to address this important workforce challenge.

Provider Interventions • INTEGRATION OF CROSS-CULTURAL EDUCATION INTO THE TRAINING OF ALL HEALTH CARE PROFESSIONALS The goal of cross-cultural education is to improve providers' ability to understand, communicate with, and care for patients from diverse backgrounds. Such education focuses on enhancing awareness of sociocultural influences on health beliefs and behaviors and on building skills to facilitate understanding and management of these factors in the medical encounter. Cross-cultural education includes curricula on health care disparities, use of interpreters, and effective communication and negotiation across cultures. These curricula can be incorporated into health professions training in medical schools, residency programs, nursing schools, and other health professions programs, and can be offered as a component of continuing education. Despite the importance of this area of education and the attention it has attracted from medical education accreditation bodies, a national survey of

senior resident physicians by Weissman and colleagues found that up to 28% felt unprepared to deal with cross-cultural issues, including caring for patients who have religious beliefs that may affect treatment, patients who use complementary medicine, patients who have health beliefs at odds with Western medicine, patients who mistrust the health care system, and new immigrants. In a study at one medical school, 70% of fourth-year students felt inadequately prepared to care for patients with limited English proficiency. Efforts to incorporate cross-cultural education into medical education will contribute to improving communication and to providing a better quality of care for all patients.

INCORPORATION OF TEACHING ON THE IMPACT OF RACE, ETHNICITY, AND CULTURE ON CLINICAL DECISION-MAKING *Unequal Treatment* and more recent studies found that stereotyping by health care providers can lead to disparate treatment based on a patient's race or ethnicity. The Liaison Committee on Medical Education, which accredits medical schools, issued a directive that medical education should include instruction on how a patient's race, ethnicity, and culture might unconsciously impact communication and clinical decision-making.

Patient Interventions Difficulty navigating the health care system and obtaining access to care can be a hindrance to all populations, particularly to minorities. Similarly, lack of empowerment or involvement in the medical encounter by minorities can be a barrier to care. Patients need to be educated on how to navigate the health care system and how best to access care. Interventions should be used to increase patients' participation in treatment decisions.

General Recommendations • INCREASE AWARENESS OF RACIAL/ETHNIC DISPARITIES IN HEALTH CARE Efforts to raise awareness of racial/ethnic health care disparities have done little for the general public but have been fairly successful among physicians, according to a Kaiser Family Foundation report. In 2006, nearly 6 in 10 people surveyed believed that blacks received the same quality of care as whites, and 5 in 10 believed that Latinos received the same quality of care as whites. These estimates are similar to findings in a 1999 survey. Despite this lack of awareness, most people believed that all Americans deserve quality care, regardless of their background. In contrast, the level of awareness among physicians has risen sharply. In 2002, the majority (69%) of physicians said that the health care system "rarely or never" treated people unfairly on the basis of their racial/ethnic background. In 2005, less than one-quarter (24%) of physicians disagreed with the statement that "minority patients generally receive lower-quality care than white patients." More recently, a survey by WebMD showed that 42% of 16,000 physicians admitted that their own personal biases impact their clinical decision-making, including on characteristics such as race and ethnicity. Increasing awareness of racial and ethnic health disparities, and their root causes, among health care professionals and the public is an important first step in addressing these disparities. The ultimate goals are to generate discourse and to mobilize action to address disparities at multiple levels, including health policymakers, health systems, and the community.

CONDUCT FURTHER RESEARCH TO IDENTIFY SOURCES OF DISPARITIES AND PROMISING INTERVENTIONS While the literature that formed the basis for the findings reported and recommendations made in *Unequal Treatment* provided significant evidence for racial and ethnic disparities, additional research is needed in several areas. First, most of the literature on disparities focuses on black-versus-white differences; much less is known about the experiences of other minority groups. Improving the ability to collect racial and ethnic patient data should facilitate this process. However, in instances where the necessary systems are not yet in place, racial and ethnic patient data may be collected prospectively in the setting of clinical or health services research to more fully elucidate disparities for other populations. Second, much of the literature on disparities to date has focused on defining areas in which these disparities exist, but less has been done to identify the multiple factors that contribute to the disparities or to test interventions to address these factors. There is clearly a need for research that identifies promising practices and solutions to disparities.

■ IMPLICATIONS FOR CLINICAL PRACTICE

Individual health care providers can do several things in the clinical encounter to address racial and ethnic disparities in health care.

Be Aware That Disparities Exist Increasing awareness of racial and ethnic disparities among health care professionals is an important first step in addressing disparities in health care. Only with greater awareness can care providers be attuned to their behavior in clinical practice and thus monitor that behavior and ensure that all patients receive the highest quality of care, regardless of race, ethnicity, or culture.

Practice Culturally Competent Care Previous efforts have been made to teach clinicians about the attitudes, values, beliefs, and behaviors of certain cultural groups—the key practice “dos and don’ts” in caring for “the Hispanic patient” or the “Asian patient,” for example. In certain situations, learning about a particular local community or cultural group, with a goal of following the principles of community-oriented primary care, can be helpful; when broadly and uncritically applied, however, this approach can actually lead to stereotyping and oversimplification of culture, without respect for its complexity.

Cultural competence has thus evolved from merely learning information and making assumptions about patients on the basis of their backgrounds to focusing on the development of skills that follow the principles of patient-centered care. *Patient-centeredness* encompasses the qualities of compassion, empathy, and responsiveness to the needs, values, and expressed preferences of the individual patient. *Cultural competence* aims to take things a step further by expanding the repertoire of knowledge and skills classically defined as “patient-centered” to include those that are especially useful in cross-cultural interactions (and that, in fact, are vital in all clinical encounters). This repertoire includes effectively using interpreter services, eliciting the patient’s understanding of his or her condition, assessing decision-making preferences and the role of family, determining the patient’s views about biomedicine versus complementary and alternative medicine, recognizing sexual and gender issues, and building trust. For example, while it is important to understand all patients’ beliefs about health, it may be particularly crucial to understand the health beliefs of patients who come from a different culture or have a different health care experience. With the individual patient as teacher, the physician can adjust his or her practice style to meet the patient’s specific needs.

Avoid Stereotyping Several strategies can allow health care providers to counteract, both systemically and individually, the normal tendency to stereotype. For example, when racially/ethnically/culturally/socially diverse teams in which each member is given equal power are assembled and are tasked to achieve a common goal, a sense of camaraderie develops and prevents the development of stereotypes based on race/ethnicity, gender, culture, or class. Thus, health care providers should aim to gain experiences working with and learning from a diverse set of colleagues. In addition, simply being aware of the operation of social cognitive factors allows providers to actively check up on or monitor their behavior. Physicians can constantly reevaluate to ensure that they are offering the same things, in the same ways, to all patients. Understanding one’s own susceptibility to stereotyping—and how disparities may result—is essential in providing equitable, high-quality care to all patients.

Work to Build Trust Patients’ mistrust of the health care system and of health care providers impacts multiple facets of the medical encounter, with effects ranging from decreased patient satisfaction to delayed care. Although the historic legacy of discrimination can never be erased, several steps can be taken to build trust with patients and to address disparities. First, providers must be aware that mistrust exists and is more prevalent among minority populations, given the history of discrimination in the United States and other countries. Second, providers must reassure patients that they come first, that everything possible will be done to ensure that they always get the best care available, and that their caregivers will serve as their advocates. Third, interpersonal skills and communication techniques that demonstrate honesty, openness, compassion, and respect on the part of the health

care provider are essential tools in dismantling mistrust. Finally, patients indicate that trust is built when there is shared, participatory decision-making and the provider makes a concerted effort to understand the patient’s background. When the doctor-patient relationship is reframed as one of solidarity, the patient’s sense of vulnerability can be transformed into one of trust. The successful elimination of disparities requires trust-building interventions and strengthening of this relationship.

■ CONCLUSION

The issue of racial and ethnic disparities in health care has gained national prominence, both with the release of the IOM report *Unequal Treatment* and with more recent articles that have confirmed their persistence and explored their root causes. Furthermore, another influential IOM report, *Crossing the Quality Chasm*, has highlighted the importance of equity—i.e., no variations in quality of care due to personal characteristics, including race and ethnicity—as a central principle of quality. Current efforts in health care reform and transformation, including a greater focus on value (high-quality care and cost-control), will sharpen the nation’s focus on the care of populations who experience low-quality, costly care. Addressing disparities will become a major focus, and there will be many obvious opportunities for interventions to eliminate them. Greater attention to addressing the root causes of disparities will improve the care provided to all patients, not just those who belong to racial and ethnic minorities.

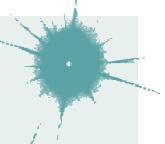
■ FURTHER READING

- BUCHMUELLER TC et al: The ACA’s impact on racial and ethnic disparities in health insurance coverage and access to care. *Health Aff* (Millwood) 39:395, 2020.
- CARNETHON MR et al: Cardiovascular health in African Americans: A scientific statement from the American Heart Association. *Circulation* 136:e393, 2017.
- DWYER-LINDGREN L et al: Inequalities in life expectancy among us counties, 1980 to 2014: Temporal trends and key drivers. *JAMA Intern Med* 177:1003, 2017.
- KREUTER MW et al: Addressing social needs in health care settings: Evidence, challenges and opportunities for public health. *Annu Rev Public Health* 42:11, 2021.
- KRIEGER N: Measures of racism, sexism, heterosexism, and gender binarism for health equity research: from structural injustice to embodied harm: An ecosocial analysis. *Annu Rev Public Health* 41:37, 2020.
- MEDSCAPE: Medscape Lifestyle Report 2016: Bias and burnout. <http://www.medscape.com/features/slideshow/lifestyle/2016/public/overview>.
- YVAS DA et al: Hidden in plain sight: Reconsidering the use of race correction in clinical algorithms. *N Engl J Med* 383:874, 2020.
- WILLIAMS DR et al: Racism and health: Evidence and needed research. *Annu Rev Public Health* 40:105, 2019.

11

Ethical Issues in Clinical Medicine

Christine Grady, Bernard Lo



Physicians face novel ethical dilemmas that can be perplexing and emotionally draining. For example, telemedicine, artificial intelligence, handheld personal devices, and learning health care systems all hold the promise of more coordinated and comprehensive care, but also raise concerns about confidentiality, the doctor-patient relationship, and responsibility. This chapter presents approaches and principles that physicians can use to address important vexing ethical issues they

encounter in their work. Physicians make ethical judgments about clinical situations every day. They should prepare for lifelong learning about ethical issues so they can respond appropriately. Traditional professional codes and ethical principles provide instructive guidance for physicians but need to be interpreted and applied to each situation. When facing or struggling with a challenging ethical issue, physicians may need to reevaluate their basic convictions, tolerate uncertainty, and maintain their integrity while respecting the opinions of others. Physicians should articulate their concerns and reasoning, discuss and listen to the views of others involved in the case, and utilize available resources, including other health care team members, palliative care, social work, and spiritual care. Moreover, ethics consultation services or a hospital ethics committee can help to clarify issues and identify strategies for resolution, including improving communication and dealing with strong or conflicting emotions. Through these efforts, physicians can gain deeper insight into the ethical issues they face and usually reach mutually acceptable resolutions to complex problems.

APPROACHES TO ETHICAL PROBLEMS

Several approaches are useful for resolving ethical issues, including approaches based on ethical principles, virtue ethics, professional oaths, and personal values. These various sources of guidance may seem to conflict in a particular case, leaving the physician in a quandary. In a diverse society, different individuals may turn to different sources of moral guidance. In addition, general moral precepts often need to be interpreted and applied to a particular clinical situation.

ETHICAL PRINCIPLES

Ethical principles can serve as general guidelines to help physicians determine the right thing to do.

Respecting Patients Physicians should always treat patients with respect, which entails understanding patients' goals, providing information, communicating effectively, obtaining informed and voluntary consent, respecting informed refusals, and protecting confidentiality. Different clinical goals and approaches are often feasible, and interventions can result in both benefit and harm. Individuals differ in how they value health and medical care and how they weigh the benefits and risks of medical interventions. Generally, physicians should respect patients' values and informed choices. Treating patients with respect is especially important when patients are responding to experiences of, or fears about, disrespect and discrimination.

GOALS AND TREATMENT DECISIONS Physicians should provide relevant and accurate information for patients about diagnoses, current clinical circumstances, expected future course, prognosis, treatment options, and uncertainties, and discuss patients' goals of care. Physicians may be tempted to withhold a serious diagnosis, misrepresent it by using ambiguous terms, or limit discussions of prognosis or risks for fear that patients will become anxious or depressed. Providing honest information about clinical situations promotes patients' autonomy and trust as well as sound communication with patients and colleagues. When physicians have to share bad news with patients, they should adjust the pace of disclosure, offer empathy and hope, provide emotional support, and call on other resources such as spiritual care or social work to help patients cope. Some patients may choose not to receive such information or may ask surrogates to make decisions on their behalf, as is common with serious diagnoses in some traditional cultures.

SHARED DECISION-MAKING AND OBTAINING INFORMED CONSENT Physicians should engage their patients in shared decision-making about their health and their care, whenever appropriate. Physicians should discuss with patients the nature, risks, and benefits of proposed care; any alternative; and the likely consequences of each option. Physicians promote shared decision-making by informing and educating patients, answering their questions, checking that they understand key issues, making recommendations, and helping them to deliberate. Medical jargon, needlessly complicated explanations, or the provision of too much information at once may overwhelm patients. Increasingly, decision aids can assist patients in playing a more active role in

decision-making, improving the accuracy of their perception of risk and benefit, and helping them feel better informed and clearer about their values. Informed consent is more than obtaining signatures on consent forms and involves disclosure of honest and understandable information to promote understanding and choice. Competent, informed patients may refuse recommended interventions and choose among reasonable alternatives. In an emergency, treatment can be given without informed consent if patients cannot give their own consent and delaying treatment while surrogates are contacted would jeopardize patients' lives or health. People are presumed to want such emergency care unless they have previously indicated otherwise.

Respect for patients does not entitle patients to insist on any care or treatment that they want. Physicians are not obligated to provide interventions that have no physiologic rationale, that have already failed, or that are contrary to evidence-based practice recommendations or good clinical judgment. Public policies and laws also dictate certain decisions—e.g., allocation of scarce medical resources during a public health crisis such as the COVID-19 pandemic, use of cadaveric organs for transplantation, and requests for physician aid in dying.

CARING FOR PATIENTS WHO LACK DECISION-MAKING CAPACITY Some patients are unable to make informed decisions because of unconsciousness, advanced dementia, delirium, or other medical conditions. Courts have the legal authority to determine that a patient is legally incompetent, but in practice, physicians usually determine when patients lack the capacity to make particular health care decisions and arrange for authorized surrogates to make decisions, without involving the courts. Patients with decision-making capacity can express a choice and appreciate their medical situation; the nature, risks, and benefits of proposed care; and the consequences of each alternative. Patient choices should be consistent with their values and not the result of delusions, hallucinations, or misinformation. Physicians should use available and validated assessment tools, resources such as psychiatry or ethics consultation, and clinical judgment to ascertain whether individuals have the capacity to make decisions for themselves. Patients should not be assumed to lack capacity if they disagree with recommendations or refuse treatment. Such decisions should be probed, however, to ensure the patient is not deciding based on misunderstandings and has the capacity to make an informed decision. When impairments are fluctuating or reversible, decisions should be postponed if possible until the patient recovers decision-making capacity.

When a patient lacks decision-making capacity, physicians seek an appropriate surrogate. Patients may designate a health care proxy through an advance directive or on a Physician Orders for Life-Sustaining Treatment form; such choices should be respected (see Chap. 12). For patients who lack decision-making capacity and have not previously designated a health care proxy, family members usually serve as surrogates. Statutes in most U.S. states delineate a prioritized list of relatives to make medical decisions. Patients' values, goals, and previously expressed preferences guide surrogate decisions. However, the patient's current best interests may sometimes justify overriding earlier preferences if an intervention is likely to provide significant benefit, previous statements do not fit the situation well, or the patient gave the surrogate leeway in decisions.

MAINTAINING CONFIDENTIALITY Maintaining confidentiality is essential to respecting patients' autonomy and privacy; it encourages patients to seek treatment and to discuss problems candidly. However, confidentiality may be overridden to prevent serious harm to third parties or the patient. Exceptions to confidentiality are justified when the risk to others is serious and probable, no less restrictive measures can avert risk, and the adverse effects of overriding confidentiality are minimized and deemed acceptable by society. For example, laws require physicians to report cases of tuberculosis, sexually transmitted infection, elder or child abuse, and domestic violence.

Beneficence or Acting in Patients' Best Interests The principle of *beneficence* requires physicians to act for the patient's benefit. Patients typically lack medical expertise, and illness may make them vulnerable. Patients rely on and trust physicians to treat them with

compassion and provide sound recommendations and treatments aimed to promote their well-being. Physicians encourage such trust and have a fiduciary duty to act in the best interests of patients, which should prevail over physicians' self-interest or the interests of third parties such as hospitals or insurers. A principle related to beneficence, "first do no harm," obliges physicians to prevent unnecessary harm by recommending interventions that maximize benefit and minimize harm and forbids physicians from providing known ineffective interventions or acting without due care. Although often cited, this precept alone provides limited guidance because many beneficial interventions also pose serious risks.

Physicians increasingly provide care within interdisciplinary teams and rely on consultation with or referral to specialists. Team members and consultants contribute different types of expertise to the provision of comprehensive, high-quality care for patients. Physicians should collaborate with and respect the contributions of the various interdisciplinary team members and should initiate and participate in regular communication and planning to avoid diffusion of responsibility and ensure accountability for quality patient care.

INFLUENCES ON PATIENTS' BEST INTERESTS Conflicts arise when patients' refusal or request of interventions thwarts their own goals for care, causes serious harm, or conflicts with their best medical interests. For example, simply accepting a young asthmatic adult's refusal of mechanical ventilation for reversible respiratory failure, in the name of respecting autonomy, is morally constricted. Physicians should elicit patients' expectations and concerns, correct their misunderstandings, and try to persuade them to accept beneficial therapies. If disagreements persist after such efforts, physicians should call on institutional resources for assistance, but patients' informed choices and views of their own best interests should prevail.

Drug prices and out-of-pocket expenses for patients have been escalating in many parts of the world and may compromise care that is in the patients' best interests. Physicians should recognize that patients, especially those with high copayments or inadequate insurance, may not be able to afford prescribed tests and interventions. Physicians should strive to prescribe medications that are affordable and acceptable to the patient. Knowing what kind of insurance, if any, the patient has and whether certain medications are likely to be covered may help in determining appropriate prescriptions. Available alternatives should be considered and discussed. Physicians should follow up with patients who don't fill prescriptions, don't take their medications, or skip doses to explore whether cost and affordability are obstacles. It may be reasonable for physicians to advocate for coverage of nonformulary products for sound reasons, such as when the formulary drugs are less effective or not tolerated or are too costly for the patient to pay for out of pocket. These should be shared decisions with the patient to the extent possible.

Organizational policies and workplace conditions may sometimes conflict with patients' best interests. Physicians' focus and dedication to the well-being and interests of patients may be negatively influenced by perceived or actual staffing inadequacies, unfair wages, infrastructural deficiencies or lack of equipment, work-hour limitations, corporate culture, and threats to personal security in the workplace. Physicians should work with institutional leaders to ensure that policies and practices support their ability to provide quality care focused on patients' best interests.

Patients' interests are served by improvements in overall quality of care and the increasing use of evidence-based practice guidelines and performance benchmarking. However, practice guideline recommendations may not serve the interests of each individual patient, especially when another plan of care may provide substantially greater benefits. In prioritizing their duty to act in the patient's best interests, physicians should be familiar with relevant practice guidelines, be able to recognize situations that might justify exceptions, and advocate for reasonable exceptions.

Acting Justly The principle of *justice* provides guidance to physicians about how to ethically treat patients and make decisions about allocating important resources, including their own time. *Justice* in a

general sense means fairness: people should receive what they deserve. In addition, it is important to act consistently in cases that are similar in ethically relevant ways, in order to avoid arbitrary, biased, and unfair decisions. Justice forbids discrimination in health care based on race, religion, gender, sexual orientation, disability, age, or other personal characteristics (**Chap. 10**).

ALLOCATION OF RESOURCES Justice also requires fair allocation of limited health care resources. Universal access to medically needed health care remains an unrealized moral aspiration in the United States and many countries around the world. Patients with no or inadequate health insurance often cannot afford health care and lack access to safety-net services. Even among insured patients, insurers may deny coverage for interventions recommended by their physician. In this situation, physicians should advocate for patients' affordable access to indicated care, try to help patients obtain needed care, and work with institutions and policies to promote wider access. Doctors might consider—or patients might request—the use of lies or deception to obtain such benefits, for example, signing a disability form for a patient who does not meet disability criteria. Although motivated by a desire to help the patient, such deception breaches basic ethical guidelines and undermines physicians' credibility and trustworthiness.

Allocation of health care resources is unavoidable when resources are limited. Allocation policies should be fair, transparent, accountable, responsive to the concerns of those affected, and proportionate to the situation, including the supply relative to the need. In the 2019–2020 SARS-CoV-2/COVID-19 pandemic, some epicenters anticipated or faced shortages of staff, protective equipment, hospital and critical care beds, and ventilators, even after increasing supplies and modifying usual clinical procedures. Many jurisdictions developed guidelines for implementing crisis standards of care to allocate limited interventions and services. Under crisis standards of care, some aspects of conventional care are not possible and interventions may not be provided to all who might benefit or wish to receive them. Crisis standards of care aim to promote the good of the community by saving the most lives in the short term, using evidence-based criteria.

When demand for medications or other interventions exceeds the supply, allocation should be fair, strive to avoid discrimination, and mitigate health disparities. First-come, first-served allocation is not fair, because it disadvantages patients who experience barriers to accessing care. To avoid discrimination, allocation decisions should not consider personal social characteristics such as race, gender, or disability, nor consider insurance status or wealth. Allocation policies also should aspire to reduce health care disparities. U.S. African-American, Latino-American, and Native-American patients suffered a disproportionate number of COVID-19 cases and deaths, likely due in part to being employed in jobs that cannot be done remotely or with physical distancing, crowded housing, lack of health benefits, and poor access to health care.

Fair and well-considered guidelines help mitigate any emotional and moral distress that clinicians may experience making difficult allocation decisions. Authorizing triage officers or committees to make allocation decisions according to policies determined with public input allows treating physicians and nurses to dedicate their efforts to their patients. Ad hoc resource allocation by physicians at the bedside may be inconsistent, unfair, and ineffective. At the bedside, physicians should act as patient advocates within constraints set by society, reasonable insurance policies, and evidence-based practice. Many allocation decisions are made at the level of public policy, with physician and public input. For example, the United Network for Organ Sharing (www.unos.org) provides criteria for allocating scarce organs.

VIRTUE ETHICS

Virtue ethics focuses on physicians' character and qualities, with the expectation that doctors will cultivate virtues such as compassion, trustworthiness, intellectual honesty, humility, and integrity. Proponents argue that, if such characteristics become ingrained, they help guide physicians in unforeseen situations. Moreover, following ethical precepts or principles without any of these virtues could lead to uncaring doctor–patient relationships.

■ PROFESSIONAL OATHS AND CODES

Professional oaths and codes are useful guides for physicians. Most physicians take oaths during their medical training, and many are members of professional societies that have professional codes. Physicians pledge to the public and to their patients that they will be guided by the principles and values in these oaths or codes and commit to the spirit of the ethical ideals and precepts represented in oaths and professional codes of ethics.

■ PERSONAL VALUES

Personal values, cultural traditions, and religious beliefs are important sources of personal morality that help physicians address ethical issues and cope with any moral distress they may experience in practice. While essential, personal morality alone is a limited ethical guide in clinical practice. Physicians have role-specific ethical obligations that go beyond their obligations as good people, including the duties to obtain informed consent and maintain confidentiality discussed earlier. Furthermore, in a culturally and religiously diverse world, physicians should expect that some patients and colleagues will have personal moral beliefs that differ from their own.

ETHICALLY COMPLEX PROFESSIONAL ISSUES FOR PHYSICIANS

■ CLAIMS OF CONSCIENCE

Some physicians, based on their personal values, have conscientious objections to providing, or referring patients for, certain treatments such as contraception or physician aid in dying. Although physicians should not be asked to violate deeply held moral beliefs or religious convictions, patients need medically appropriate, timely care and should always be treated with respect. Institutions such as clinics and hospitals have a collective ethical duty to provide care that patients need while making reasonable attempts to accommodate health care workers' conscientious objections—for example, when possible by arranging for another professional to provide the service in question. Patients seeking a relationship with a doctor or health care institution should be notified in advance of any conscientious objections to the provision of specific interventions. Since insurance often constrains patients' selection of physicians or health care facilities, switching providers can be burdensome. There are also important limits on claims of conscience. Health care workers may not insist that patients receive unwanted medical interventions. They also may not refuse to treat or discriminate against patients because of their race, ethnicity, disability, genetic information, or diagnosis. Such discrimination is illegal and violates physicians' duties to respect patients. Refusal to treat patients for other reasons such as sexual orientation, gender identity, or other personal characteristics is legally more controversial, yet ethically inappropriate because it falls short of helping patients in need and respecting them as persons.

■ PHYSICIAN AS GATEKEEPER

In some cases, patients may ask their physicians to facilitate access to services that the physician has ethical qualms about providing. For example, a patient might request a prescription for a cognitively enhancing medication to temporarily augment his cognitive abilities in order to take an exam or apply for employment. Patients may request more pain medication than the physician believes is warranted for the given situation or marijuana to facilitate sleep. Patients may ask their physician to sign a waiver to avoid vaccines for reasons that are not included in state exceptions (see *Chap. 3*). A physician may feel uncomfortable prescribing attention-deficit/hyperactivity disorder medications to a young child because she is not convinced that the possible benefit justifies the risks to the child despite the parent's request. In these circumstances, the physician should work with the patient or parent to understand the reasons for their requests, some of which might be legitimate. In addition to considering possible risks and benefits to the patient, the physician should consider how meeting the request might affect other patients, societal values, and public trust in the medical profession. If the physician determines that fulfilling the request requires deception, is unfair, jeopardizes her professional

responsibilities, or is inconsistent with the patient's best medical interests, the physician should decline and explain the reasons to the patient.

■ MORAL DISTRESS

Health care providers, including residents, medical students, and experienced physicians, may experience moral distress when they feel that ethically appropriate action is hindered by institutional policies or culture, decision-making hierarchies, limited resources, or other reasons. Moral distress can lead to anger, anxiety, depression, frustration, fatigue, work dissatisfaction, and burnout. A physician's health and well-being can affect how he or she cares for patients. Discussing complex or unfamiliar clinical situations with colleagues and seeking assistance with difficult decisions can help alleviate moral distress, as can a healthy work environment characterized by open communication, mutual respect, and emphasis on the common goal of good patient care. In addition, physicians should take good care of their own well-being and be aware of the personal and system factors associated with stress, burnout, and depression. Health care organizations should provide a supportive work environment, counseling, and other support services when needed.

■ OCCUPATIONAL RISKS AND BURDENS

Physicians accept some physical risk in fulfilling their professional responsibilities, including exposure to infectious agents or toxic substances, violence in the workplace, and musculoskeletal injury. Nonetheless, most physicians, nurses, and other hospital staff willingly care for patients, despite personal risk and fear, grueling hours, and sometimes inadequate personal protective equipment or information. During the COVID-19 pandemic, many communities honored clinicians' dedication to professional ideals, and some medical students who were relieved from in-person patient care responsibilities volunteered to support front-line workers in other ways. The burdens of navigating professional and personal responsibilities fall more heavily on women health care providers. Health care institutions are responsible for reducing occupational risk and burden by providing proper information, training and supervision, protective equipment, infrastructure and workflow modifications, and emotional and psychological support to physicians. Clinical leaders need to acknowledge fears about personal safety and take steps to mitigate the impact of work on family responsibilities, moral distress, and burnout.

■ USE OF SOCIAL MEDIA AND PATIENT PORTALS

Increasingly, physicians use social and electronic media to share information and advice with patients and other providers. Social networking may be especially useful in reaching young or otherwise hard-to-access patients. Patients increasingly access their physicians' notes through patient portals, which aim to transparently share information, promote patient engagement, and increase adherence. Physicians should be professional and respectful and consider patient confidentiality, professional boundaries, and therapeutic relationships when posting to social media or writing notes for the portal. Overall, appropriate use of these platforms can enhance communication and transparency while avoiding misunderstandings or harmful consequences for patients, physicians, or their colleagues. Unprofessional or careless posts that express frustration or anger over work incidents, disparage patients or colleagues, use offensive or discriminatory language, or reveal inappropriate personal information about the physician can have negative consequences. Physicians should separate professional from personal websites and accounts and follow institutional and professional society guidelines when communicating with patients.

CONFLICTS OF INTEREST

Acting in patients' best interests may sometimes conflict with a physician's self-interest or the interests of third parties such as insurers or hospitals. From an ethical viewpoint, patients' interests are paramount. Transparency, appropriate disclosure, and management of conflicts of interest are essential to maintain the trust of colleagues and the public. Disclosure requirements vary for different purposes, and software has

been developed to assist physicians in complying with specific requirements. Importantly, not all conflicts are financial. Physicians sometimes face conflicts of commitment between their patient's interests and their own personal interests, professional goals, responsibilities, and aspirations. As mentioned earlier, physicians should prioritize patients' interests while recognizing possible conflicts and using disclosure, discussion with the chief of service, and management of the conflict or recusal when appropriate.

In addition to individual physicians, medical institutions may have conflicts of interest arising from patent rights, industry-funded research programs, and donations from individuals and companies. Institutions need to be transparent about the presence and amount of such relationships and make clear the steps taken to prevent such relationships from having an impact on clinical or financial decisions. If there is good evidence that a donor acted in ways that breached ethical or legal standards, the institution should take steps not to benefit from the donation or honor the donor.

■ FINANCIAL INCENTIVES

Physicians have financial incentives to improve the quality or efficiency of care that might lead some to avoid patients who are older, are chronically ill, or have more complicated problems, or to focus on benchmarked outcomes even when not in the best interests of individual patients. In contrast, fee-for-service payments might encourage physicians to order more interventions than necessary or to refer patients to laboratory, imaging, or surgical facilities in which they have a financial stake. Regardless of financial incentives, physicians should recommend available care that is in the patient's best interests—no more and no less.

■ RELATIONSHIPS WITH PHARMACEUTICAL COMPANIES

Financial relationships between physicians and industry are increasingly scrutinized. Many academic medical centers have banned drug-company gifts, including branded pens and notepads and meals to physicians, to reduce inappropriate risk of undue influence or subconscious feelings of reciprocity and to decrease possible influences on public trust or the costs of health care.

The federal Open Payments website provides public information on the payments and amounts that drug and device companies give to individual physicians by name. The challenge is to distinguish payments for scientific consulting and research contracts—which should be encouraged as consistent with professional and academic missions—from those for promotional speaking and consulting whose goal is to increase sales of company products.

■ LEARNING CLINICAL SKILLS

Medical students', residents', and physicians' interests in learning, which fosters the long-term goal of benefiting future patients, may sometimes conflict with the short-term goal of providing optimal care to current patients. When trainees are learning procedures on patients, they lack the proficiency of experienced physicians, and patients may experience inconvenience, discomfort, longer procedures, or increased risk. Increasingly, institutions are developing clinical skills laboratories for simulation-based medical education and requiring students to demonstrate proficiency before carrying out procedures such as venipuncture and intravenous lines in patients. Furthermore, teaching hospitals are establishing proceduralist services in which procedure-specialist faculty members directly supervise interns for procedures such as lumbar puncture and thoracentesis and certify their proficiency. Medical students may need to defer learning such invasive procedures until internship. Seeking patients' consent for trainee participation in their care is always important and is particularly important for intimate examinations, such as pelvic, rectal, breast, and testicular examinations, and for invasive procedures. Patients should be told who is providing care and how trainees are supervised. Failing to introduce students or not telling patients that trainees will be performing procedures undermines trust, may lead to more elaborate deception, and makes it difficult for patients to make informed choices about their care. Most patients, when informed, allow trainees to play an active role in their care.

■ RESPONSE TO MEDICAL ERRORS

Errors are inevitable in clinical medicine, and some errors cause harm to patients. Most errors are caused by lapses of attention or flaws in the system of delivering health care; only a small number result from blameworthy individual behavior. Many health care institutions have adopted a just culture system, which encourages open and honest reporting of errors as essential to quality learning and shifts the focus from individual blame to system design for improvement in quality and safety (**Chap. 8**). This approach is more likely than a punitive approach to improve patient safety. However, professional discipline is appropriate for cases of gross incompetence, reckless behavior, physician impairment, and boundary violations. Physicians and students may fear that disclosing errors will damage their careers. Physicians and health care institutions show respect for patients by disclosing and explaining errors, offering an apology, offering appropriate compensation for harm done, and using errors as opportunities to improve the quality of care.

■ PHYSICIAN IMPAIRMENT

Physicians may hesitate to intervene when colleagues impaired by alcohol, drugs, or psychiatric or medical illness place patients at risk. However, society relies on physicians to regulate themselves. Colleagues of an impaired physician should take steps to protect patients and help their impaired colleague, starting with reporting their concerns to their clinical supervisor or director.

ETHICAL ISSUES IN CLINICAL RESEARCH

Clinical research is essential to translate scientific discoveries into beneficial interventions for patients. However, clinical research raises ethical concerns because participants face inconvenience and risks in research designed to advance scientific knowledge and not specifically to benefit them. Ethical guidelines require researchers to rigorously design and conduct research, minimize risk to participants, and obtain informed and voluntary consent from participants and approval from an institutional review board (IRB). IRBs determine that risks to participants are acceptable and have been minimized and recommend appropriate additional protections when research includes vulnerable participants.

Physicians may be clinical research investigators themselves or may be in a position to refer or recommend clinical trial participation to their patients. Physician-investigators are likely to feel some inherent tension between conducting research and providing health care. Awareness of this tension, familiarity with research ethics, collaboration with research and clinical team members, and utilizing research ethics consultation can help to mitigate tensions. Before starting clinical research, investigators should complete training in the ethics of clinical research, which is widely available.

Physicians also should be critical consumers of clinical research results and keep up with research advances that change standards of practice. Precision medicine initiatives aim to individualize clinical care by combining clinical information from electronic health records, genomic sequencing, and data from personal mobile devices. Furthermore, physicians and health care institutions are analyzing data routinely collected and available in electronic health records, leftover clinical specimens, and administrative data. Such studies encompass traditional discovery research as well as quality improvement, comparative effectiveness research, and learning health care systems. Efforts to improve the quality of care in real-world clinical settings are important but also raise new issues about informed consent, privacy, and risk.

EMERGING TECHNOLOGIES

Scientific advances in genome sequencing, gene editing (e.g., with CRISPR-Cas9), machine learning, artificial intelligence, computer-brain interfaces, and other technologies offer great promise for research and clinical care with the ultimate goal of improving the prediction, prevention, and treatment of disease. Groundbreaking innovations that have strong scientific plausibility need to be evaluated in rigorous clinical studies for efficacy and safety.

Physicians should keep up to date on the status of novel and often complex technologies as research evolves, data emerge, and technologies are incorporated into clinical practice. They can help their patients understand research findings and the evidence for clinical use, correct any misunderstandings, facilitate shared decision-making, and advocate for fair access to such therapies. Further, physicians should engage in professional and public discussion related to allocation of resources and fair access to expensive new therapies and emerging technologies and their impact on overall health care affordability.

Certain cell-based therapies, such as peripheral blood stem cell transplantation (**Chap. 114**) and chimeric antigen receptor (CAR)-T cell therapy (**Chap. 69**), are approved for use in several serious hematologic cancers, and gene therapies have been approved as safe and effective for clinical use in certain serious inherited diseases and cancers. Patients may request these and other complex, highly technical, and expensive therapies for unproven indications. Yet, claims of cures through unproven stem cell or gene-based “therapies” pose significant health and financial risks to patients without evidence of benefit. Physicians should help patients distinguish approved therapies from unproven claims and refer interested patients to well-designed clinical trials.

Medical applications of CRISPR-Cas9 are promising, and their safety and efficacy for particular clinical conditions are being carefully evaluated in clinical trials. Applications of CRISPR genome editing in somatic cells to modify or correct problematic genes could lay the foundation for treating a variety of serious diseases, including blood disorders, HIV, cancer, and hereditary blindness. Germline gene editing in blastocysts or embryos raises many ethical questions and is currently not permitted in the United States in clinical trials or clinical practice.

In artificial intelligence (AI), computers carry out tasks typically done by humans. Machine learning (ML) is a type of AI that automatically learns and improves its performance without explicit programming. Clinical algorithms using AI and ML can make diagnoses from radiology images, retinal scans, or skin photographs and identify patients at increased risk for surgical complications, critical care, or hospital readmission. However, such algorithms can also pose risks. Bias may occur if an algorithm was derived or validated from a data set in which groups who suffer from health disparities or poor health outcomes are underrepresented or if the algorithm predicts outcomes that are not clinically meaningful. To address these ethical concerns, researchers should assess AI algorithms in well-designed randomized clinical trials with clinical endpoints. Institutions should integrate validated and unbiased algorithms into clinical workflow without unduly burdening physicians and nurses and should check effectiveness and safety in their particular settings and patient populations.

Physicians should stay informed of emerging evidence about such technologies and the ethical challenges that accompany their use and always keep their patients’ best interests and preferences at the forefront.

GLOBAL CONSIDERATIONS

■ INTERNATIONAL RESEARCH

Clinical research is often conducted across multiple sites and across national borders. Societal, legal, and cultural norms and perspectives about research may vary, and there are many ethical challenges. Physician-investigators involved in international research should be familiar with international guidelines, such as the Declaration of Helsinki, the Council for International Organizations of Medical Sciences (CIOMS) guidelines, and the International Council on Harmonisation Good Clinical Practice guidelines, as well as national and local laws where research is taking place. Partnering with local researchers and communities is essential not only to demonstrate respect but also to facilitate successful clinical research.

■ INTERNATIONAL CLINICAL EXPERIENCES

Many physicians and trainees gain valuable experience providing patient care in international settings through international training opportunities or volunteering for humanitarian or other international

clinical work. Such arrangements, however, raise ethical challenges—for example, as a result of differences in beliefs about health and illness, expectations regarding health care and physicians’ roles, standards of clinical practice, resource limitations, and norms for disclosure of serious diagnoses. Additional dilemmas arise if visiting physicians and trainees take on responsibilities beyond their expertise or if donated drugs and equipment are not appropriate to local needs. Visiting physicians and trainees should prepare well for these experiences, receive training and mentoring, learn about cultural and clinical practices in the host community, respect local customs and values, collaborate closely with local professionals and staff, and be explicit and humble about their own skills, knowledge, and limits. Leaders of global health field experiences should ensure that participating physicians receive training on ethical and cultural issues, as well as mentoring, backup, and debriefing upon return home.

■ CONCLUSION

Ethical issues are common in clinical medicine and occur in circumstances that may be foreseeable, novel, or unexpected. Physicians address these ethical issues by being prepared, informed, and thoughtful and using appropriate available resources.

■ FURTHER READING

- BEAUCHAMP T, CHILDRESS J: *Principles of Biomedical Ethics*, 8th ed. New York, Oxford University Press, 2019.
- DEJONG C et al: An ethical framework for allocating scarce medications for COVID-19 in the US. *JAMA* 323:2367, 2020.
- MATHENY M et al (eds): *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. NAM Special Publication. Washington, DC, National Academy of Medicine, 2019.
- ULRICH C, GRADY C: *Moral Distress in the Health Professions*. Cham, Switzerland, Springer-Nature International, 2018.
- WASSERMAN J et al: Responding to unprofessional behavior by trainees: a “just culture” framework. *N Engl J Med* 382:773, 2020.
- WICCLAIR MR: Conscientious objection, moral integrity, and professional obligations. *Perspect Biol Med* 62:543, 2019.

12

Palliative and End-of-Life Care

Ezekiel J. Emanuel



EPIDEMIOLOGY

■ CAUSES OF DEATH

In 2019, 2,854,838 individuals died in the United States (**Table 12-1**). Approximately 74% of these deaths occurred in those aged ≥65 years. The epidemiology of death has changed significantly since 1900 and even since 1980. In 1900, heart disease caused ~8% of all deaths, and cancer accounted for <4% of all deaths. In 1980, heart disease accounted for 38.2% of all deaths, cancer 20.9%, and cerebrovascular disease 8.6% of all deaths. By 2019, there had been a dramatic drop in deaths from cardiovascular and cerebrovascular diseases. In 2019, 23.1% of all deaths were from cardiovascular disease and just 5.3% from cerebrovascular disease. Deaths attributable to cancer, however, had increased slightly to 21.0%. The proportions of deaths due to chronic lower respiratory disease, diabetes, Alzheimer’s, and suicide have increased. Interestingly, in 2019, HIV/AIDS accounted for <0.18% of all U.S. deaths. While unlikely to continue being a leading cause of death in the future, COVID-19 was also the cause for >600,000 deaths in 2020–2021, and the official figure is almost certainly an undercount of the actual death toll.

TABLE 12-1 Ten Leading Causes of Death in the United States and Britain

CAUSE OF DEATH	UNITED STATES (2019)		ENGLAND AND WALES (2019)	
	NUMBER OF DEATHS, ALL AGES (%)	NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE	NUMBER OF DEATHS, ALL AGES (%)	NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE
All deaths	2,854,838	2,117,332	530,841	449,047
Heart disease ^a	659,041 (23.1)	531,583 (25.1)	87,095 (16.4)	74,967 (16.7)
Malignant neoplasms	599,601 (21.0)	435,462 (20.6)	147,419 (27.8)	118,982 (26.5)
Chronic lower respiratory diseases	156,979 (5.5)	133,246 (6.3)	31,221 (5.9)	28,235 (6.3)
Accidents	173,040 (6.1)	60,527 (2.9)	15,141 (2.9)	8999 (2.0)
Cerebrovascular diseases	150,005 (5.3)	129,193 (6.1)	29,816 (5.6)	27,210 (6.0)
Alzheimer's disease	121,499 (4.3)	120,090 (5.7)	20,400 (3.8)	20,279 (4.5)
Diabetes mellitus	87,847 (3.1)	62,397 (2.9)	6528 (1.2)	5552 (1.2)
Influenza and pneumonia	49,783 (1.7)	40,399 (1.9)	26,398 (5.0)	24,269 (5.4)
Nephritis, nephritic syndrome, nephrosis	51,565 (1.8)	42,230 (2.0)	3575 (0.7)	3323 (0.7)
Intentional self-harm	47,511 (1.7)	—	4832 (0.9)	751 (0.2)

^aCalculated using International Classification of Diseases codes I00–I09, I11, I13, I20–I51.

Source: National Center for Health Statistics (United States, 2019), <http://www.cdc.gov/nchs>; National Statistics (Great Britain, 2019), <http://www.statistics.gov.uk>.

This change in the epidemiology of death is also reflected in the costs of illness. In the United States, ~84% of all health care spending goes to patients with chronic illnesses, and 12% of total personal health care spending—slightly less than \$400 billion in 2015—goes to the 0.83% of the population in the last year of their lives.

In upper-middle- and upper-income countries, an estimated 70% of all deaths are preceded by a disease or condition, making it reasonable to plan for dying in the foreseeable future. Cancer has served as the paradigm for terminal care, but it is not the only type of illness with a recognizable and predictable terminal phase. Since heart failure, chronic obstructive pulmonary disease (COPD), chronic liver failure, dementia, and many other conditions have recognizable terminal phases, a systematic approach to end-of-life care should be part of all medical specialties. Many patients with chronic illness-related symptoms and suffering also can benefit from palliative care regardless of prognosis. Ideally, palliative care should be considered part of comprehensive care for all chronically ill patients. Strong evidence demonstrates that palliative care can be improved by coordination between caregivers, doctors, and patients for advance care planning, as well as dedicated teams of physicians, nurses, and other providers.

SITE OF DEATH

Where patients die varies by country. In Belgium and Canada, for instance, over half of all cancer patients still die in the hospital. The past few decades have seen a steady shift, both in the United States and other countries like the Netherlands, out of the hospital, as patients and their families list their own homes as the preferred site of death. In the early 1980s, ~70% of American cancer patients died in the hospital. Today, that percentage is ~25% (Fig. 12-1). A recent report shows that since 2000, there has been a shift in the United States from inpatient to home deaths, especially for patients with cancer, COPD, and dementia. For instance, among Medicare beneficiaries, 30.1% of deaths due to cancer in 2000 occurred in acute care hospitals; by 2009, this figure had dropped to 22.1%; by 2015, it was 19.8%.

Paradoxically, while deaths in acute care hospitals have declined in the United States since 2000, both hospitalizations in the last 90 days of life and—even more troublingly—admission to the intensive care unit (ICU) in the last 30 days have actually increased. Over 40% of cancer patients in the United States are admitted to the ICU in their last 6 months of life, and >25% of cancer patients are admitted to the hospital in the last 30 days.

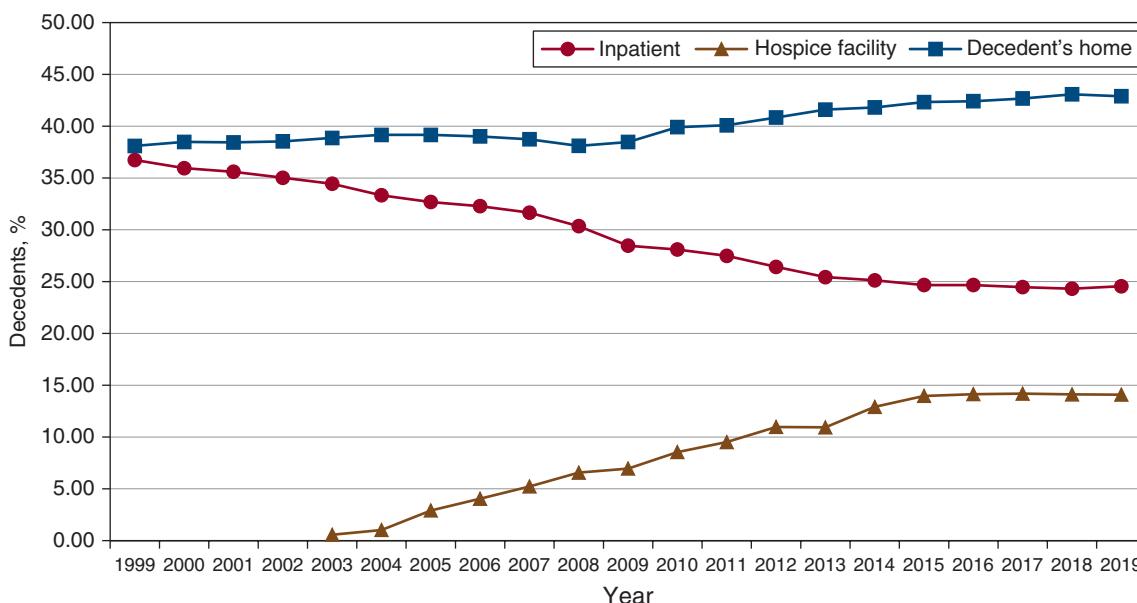


FIGURE 12-1 Graph showing trends in cancer decedents' site of death 1999–2019. (Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999–2019 on CDC WONDER Online Database. <http://wonder.cdc.gov>.)

The shift in deaths out of the hospital has been accompanied by an increase in the use of hospice in the United States. In 2000, 21.6% of Medicare decedents used hospice at the time of death; by 2009, 42.2% were using hospice; and by 2018, 50.7% of Medicare decedents were enrolled in hospice at the time of death. Among cancer patients, ~60% were using hospice at the time of death. Hospice is also increasingly being used by noncancer patients. Today, cancer patients constitute ~20% of hospice users. But since 2014, the proportion of patients with other diagnoses using hospice has grown substantially, including those with circulatory/heart disease (17.4% in 2018 vs 13.8% in 2014), stroke (9.5% vs 6.2%), and respiratory disease (11.0% vs 9.4%). Of 2018 Medicare hospice decedents, 51.5% died at home, 17.4% in a nursing facility, 12.8% in a hospice inpatient facility, and 12.3% in assisted living.

Unfortunately, significant racial disparities exist in end-of-life care and the use of hospice, especially for noncancer deaths. Racial and ethnic minorities are less likely to receive hospice services than white decedents and are more likely to receive invasive or aggressive care in end-of-life treatment. Of people who died of head and neck cancers between 1999 and 2017, African Americans and Asians/Pacific Islanders were less likely to die at home or in hospice. Among Medicare beneficiaries who had a pancreatectomy for pancreatic cancer and lived at least 30 days, racial and ethnic minority patients remained 22% less likely than white patients to initiate hospice before death.

In 2008, for the first time, the American Board of Medical Specialties (ABMS) offered certification in hospice and palliative medicine. With the shortening of hospital stays, many serious conditions are now being treated at home or on an outpatient basis. Consequently, providing optimal palliative and end-of-life care requires ensuring that appropriate services are available in a variety of settings, including noninstitutional settings.

HOSPICE AND THE PALLIATIVE CARE FRAMEWORK

Central to this type of care is an interdisciplinary team approach that typically encompasses pain and symptom management, spiritual and psychological care for the patient, and support for family caregivers during the patient's illness and the bereavement period.

One of the more important changes in this field is beginning palliative care many months before death in order to focus on symptom relief and then switching to hospice in the patient's last few months. This approach avoids leaving hospice until the very end by introducing palliative care earlier, thereby allowing patients and families time to accommodate and transition. Phasing palliative care into end-of-life care means that patients will often receive palliative interventions long before they are formally diagnosed as terminally ill, or likely to die within 6 months.

Fundamental to ensuring quality palliative and end-of-life care is a focus on four broad domains: (1) physical symptoms; (2) psychological symptoms; (3) social needs that include interpersonal relationships, caregiving, and economic concerns; and (4) existential or spiritual needs.

ASSESSMENT AND CARE PLANNING

Comprehensive Assessment Standardized methods for conducting a comprehensive assessment focus on evaluating the patient's condition in all four domains affected by the illness: physical, psychological, social, and spiritual.

A comprehensive assessment should follow a modified version of the traditional medical history and physical examination and should emphasize both physical and mental symptoms. Questions should aim to elucidate symptoms, discern sources of suffering, and gauge how much those symptoms interfere with the patient's quality of life. Standardized and repeated assessments to evaluate the effectiveness of interventions are critical. Thus, clinicians should use shorter, validated instruments, such as (1) the revised Edmonton Symptom Assessment Scale; (2) Condensed Memorial Symptom Assessment Scale (MSAS); (3) MD Anderson Brief Symptom Inventory; (4) Rotterdam Symptom Checklist; (5) Symptom Distress Scale; (6) Patient-Reported Outcomes Measurement Information System; and (7) Interactive Symptom Assessment and Collection (ISAAC) tool.

MENTAL HEALTH With respect to mental health, many practices use the Patient Health Questionnaire-9 (PHQ-9) to screen for depression and the Generalized Anxiety Disorder-7 (GAD-7) to screen for anxiety. Using such tools ensures that the assessment is comprehensive and does not focus excessively on only pain.

INVASIVE TESTS Invasive tests are best avoided in end-of-life care, and even minimally invasive tests should be evaluated carefully for their benefit-to-burden ratio for the patient. Aspects of the physical examination that are uncomfortable and unlikely to yield useful information that change patient management should be omitted.

SOCIAL NEEDS Health care providers should also assess the status of important relationships, financial burdens, caregiving needs, and access to medical care. Relevant questions include the following: *How often is there someone to feel close to? How has this illness been for your family? How has it affected your relationships? How much help do you need with things like getting meals and getting around? How much trouble do you have getting the medical care you need?*

EXISTENTIAL NEEDS To determine a patient's existential needs, providers should assess distress, the patient's sense of emotional and existential well-being, and whether the patient believes he or she has found purpose or meaning. Helpful assessment questions can include the following: *How much are you able to find meaning since your illness began? What things are most important to you at this stage?*

PERCEPTION OF CARE In addition, it can be helpful to ask how the patient perceives his or her care: *How much do you feel your doctors and nurses respect you? How clear is the information from us about what to expect regarding your illness? How much do you feel that the medical care you are getting fits with your goals?* If concern is detected in any of these areas, deeper evaluative questions are warranted.

Communication Particularly when an illness is life-threatening, there exists the potential for many emotionally charged and potentially conflict-creating moments—collectively called “bad news” situations—in which empathic and effective communication skills are essential. Those moments include the sharing of a terminal diagnosis with the patient and/or family, the discussion of the patient's prognosis and any treatment failures, the consideration of deemphasizing efforts to cure and prolong life while focusing more on symptom management and palliation, advance care planning, and the patient's actual death. Although these conversations can be difficult, research indicates that end-of-life discussions can lead to earlier hospice referrals, rather than overly aggressive treatment, ultimately benefiting quality of life for patients and improving the bereavement process for families.

Just as surgeons prepare for major operations and investigators rehearse a presentation of research results, physicians and health care providers caring for patients with significant or advanced illnesses should develop a standardized approach for sharing important information and planning interventions. In addition, physicians must be aware that families often care not only about how prepared the physician was to deliver bad news, but also the setting in which it was delivered. For instance, one study found that 27% of families making critical decisions for patients in an ICU desired better and more private physical space to communicate with physicians.

One structured seven-step procedure for communicating bad news goes by the acronym P-SPIKES: (1) *prepare* for the discussion, (2) *set up* a suitable environment, (3) *begin* the discussion by finding out what the patient and/or family understand, (4) *determine* how they will comprehend new information best and how much they want to know, (5) *provide* needed new knowledge accordingly, (6) *allow* for emotional responses, and (7) *share* plans for the next steps in care (Table 12-2).

Continuous Goal Assessment Major barriers to providing high-quality palliative and end-of-life care include the difficulty in determining an accurate prognosis and the emotional resistance of patients and their families to accepting the implications of a poor prognosis. A practical solution to these barriers is to integrate palliative care interventions or home visits from a palliative care visiting nurse months before the estimated final 6 months of life. Under this

TABLE 12-2 Elements of Communicating Bad News—The P-SPIKES Approach

ACRONYM	STEPS	AIM OF THE INTERACTION	PREPARATIONS, QUESTIONS, OR PHRASES
P	Preparation	Mentally prepare for the interaction with the patient and/or family.	Review what information needs to be communicated. Plan how you will provide emotional support. Rehearse key steps and phrases in the interaction.
S	Setting of the interaction	Ensure the appropriate setting for a serious and potentially emotionally charged discussion.	Ensure that patient, family, and appropriate social supports are present. Devote sufficient time. Ensure privacy and prevent interruptions by people or beeper. Bring a box of tissues.
P	Patient's perception and preparation	Begin the discussion by establishing the baseline and whether the patient and family can grasp the information. Ease tension by having the patient and family contribute.	Start with open-ended questions to encourage participation. Possible questions to use: <i>What do you understand about your illness?</i> <i>When you first had symptom X, what did you think it might be?</i> <i>What did Dr. X tell you when he or she sent you here?</i> <i>What do you think is going to happen?</i>
I	Invitation and information needs	Discover what information needs the patient and/or family have and what limits they want regarding the bad information.	Possible questions to use: <i>If this condition turns out to be something serious, do you want to know?</i> <i>Would you like me to tell you all the details of your condition? If not, who would you like me to talk to?</i>
K	Knowledge of the condition	Provide the bad news or other information to the patient and/or family sensitively.	Do not just dump the information on the patient and family. Check for patient and family understanding. Possible phrases to use: <i>I feel badly to have to tell you this, but...</i> <i>Unfortunately, the tests showed...</i> <i>I'm afraid the news is not good...</i>
E	Empathy and exploration	Identify the cause of the emotions—e.g., poor prognosis. Empathize with the patient's and/or family's feelings. Explore by asking open-ended questions.	Strong feelings in reaction to bad news are normal. Acknowledge what the patient and family are feeling. Remind them such feelings are normal, even if frightening. Give them time to respond. Remind the patient and family you won't abandon them. Possible phrases to use: <i>I imagine this is very hard for you to hear.</i> <i>You look very upset. Tell me how you are feeling.</i> <i>I wish the news were different.</i> <i>We'll do whatever we can to help you.</i>
S	Summary and planning	Delineate for the patient and the family the next steps, including additional tests or interventions.	It is the unknown and uncertain that can increase anxiety. Recommend a schedule with goals and landmarks. Provide your rationale for the patient and/or family to accept (or reject). If the patient and/or family are not ready to discuss the next steps, schedule a follow-up visit.

Source: Adapted from R Buckman: *How to Break Bad News: A Guide for Health Care Professionals*. Baltimore, Johns Hopkins University Press, 1992.

approach, palliative care no longer conveys the message of failure, having no more treatments, or “giving up hope.” The transition from palliative to end-of-life care or hospice also feels less hasty and unexpected to the family. Fundamental to integrating palliative care with curative therapy is the inclusion of a continuous goal assessment as part of the routine patient reassessments that occur at most patient-physician encounters.

Goals for care are numerous, ranging from curing a specific disease, to prolonging life, to relieving a particular symptom, to adapting to a progressive disability without disrupting the family, to finding peace of mind or personal meaning, to dying in a manner that leaves loved ones with positive memories. Discerning a patient’s goals for care can be approached through a seven-step protocol: (1) ensure that medical and other information is as complete as reasonably possible and is understood by all relevant parties (see above); (2) explore what the patient and/or family is hoping for, while also identifying relevant and realistic goals; (3) share all the options with the patient and family; (4) respond with empathy as they adjust to changing expectations; (5) make a plan that emphasizes what can be done to achieve the realistic goals; (6) follow through with the plan; and (7) periodically review the plan and consider at every encounter whether the goals of care should be revised with the patient and/or family. Each of these steps need not be followed in rote order, but together they provide a helpful framework for interactions with patients and their families regarding

their goals for care. Such interactions can be especially challenging if a patient or family member has difficulty letting go of an unrealistic goal. In such cases, the provider should help them refocus on more realistic goals and should also suggest that while it is fine to hope for the best, it is still prudent to plan for other outcomes as well.

Advance Care Planning • PRACTICES Advance care planning is the process of planning for future medical care in case the patient becomes incapable of making medical decisions. A 2010 study of adults aged ≥60 who died between 2000 and 2006 found that while 42% of adults were required to make treatment decisions in their final days of life, 70% lacked decision-making capacity. Among those lacking decision-making capacity, approximately one-third did not have advance planning directives. Ideally, such planning would occur before a health care crisis or the terminal phase of an illness. Unfortunately, diverse barriers prevent this. Approximately 80% of Americans endorse advance care planning and living wills. However, according to a 2013 Pew survey, only 35% of adults have written down their end-of-life wishes. Other studies report that even fewer Americans—with some estimates as low as 26% of adults—have filled out advance care directives. A review of studies suggests that the percentage of Americans who had written advance directives did not change between 2011 and 2016 and remains slightly over one-third of Americans. Larger numbers of adults, between 50 and 70%, claim to have talked with someone

about their treatment wishes. Americans aged 65 and older are more likely to complete an advance directive compared to younger adults (46% vs 32%).

Effective advance care planning should follow six key steps: (1) introducing the topic, (2) structuring a discussion, (3) reviewing plans that have been discussed by the patient and family, (4) documenting the plans, (5) updating them periodically, and (6) implementing the advance care directives (**Table 12-3**). Two of the main barriers to advance care planning are problems in raising the topic and difficulty in structuring a succinct discussion. Raising the topic can be done efficiently as a routine matter, noting that it is recommended for all patients, analogous to purchasing insurance or estate planning. Many of the most difficult cases have involved unexpected, acute episodes of brain damage in young individuals.

Structuring a focused discussion is an important communication skill. To do so, a provider must first identify the health care proxy and recommend his or her involvement in the advance care planning process. Next, a worksheet must be selected that has been demonstrated to produce reliable and valid expressions of patient preferences, and the patient and proxy must be oriented to it. Such worksheets exist for both general and disease-specific situations. The provider should then discuss with the patient and proxy one example scenario to demonstrate how to think about the issues. It is often helpful to begin with a scenario in which the patient is likely to have settled preferences for care, such as being in a persistent vegetative state. Once the patient's preferences

for interventions in this scenario are determined, the provider should suggest that the patient and proxy discuss and complete the worksheet for each other. If appropriate, the patient and proxy should consider involving other family members in the discussion. During a subsequent return visit, the provider should go over the patient's preferences, checking and resolving any inconsistencies. After having the patient and proxy sign the document, the provider should place the document in the patient's medical chart and make sure that copies are provided to relevant family members and care sites. Since patients' preferences can change, these documents must be reviewed periodically.

TYPES OF DOCUMENTS Advance care planning documents are of two broad types. The first includes living wills, also known as instructional directives; these are advisory documents that describe the types of decisions that should direct a patient's care. Some are more specific, delineating different scenarios and interventions for the patient to choose from. Among these, some are for general use and others are designed for use by patients with a specific type of disease, such as cancer, renal failure, or HIV. Less specific directives can be general statements, such as not wanting life-sustaining interventions, or forms that describe the values that should guide specific discussions about terminal care. The second type of advance directive allows the designation of a health care proxy (sometimes also referred to as a durable attorney for health care), an individual selected by the patient to make decisions. The choice is not either/or; a combined directive that includes a living

TABLE 12-3 Steps in Advance Care Planning

STEP	GOALS TO BE ACHIEVED AND MEASURES TO COVER	USEFUL PHRASES OR POINTS TO MAKE
Introduce advance care planning	Ask the patient what he or she knows about advance care planning and if he or she has already completed an advance care directive.	<i>I'd like to talk with you about something I try to discuss with all my patients. It's called advance care planning. In fact, I feel that this is such an important topic that I have done this myself. Are you familiar with advance care planning or living wills?</i>
	Indicate that you as a physician have completed advance care planning.	<i>Have you thought about the type of care you would want if you ever became too sick to speak for yourself? That is the purpose of advance care planning.</i>
	Indicate that you try to perform advance care planning with all patients regardless of prognosis.	<i>There is no change in health that we have not discussed. I am bringing this up now because it is sensible for everyone, no matter how well or ill, old or young.</i>
	Explain the goals of the process as empowering the patient and ensuring that you and the proxy understand the patient's preferences.	<i>Have many copies of advance care directives available, including in the waiting room, for patients and families.</i>
	Provide the patient relevant literature, including the advance care directive that you prefer to use.	<i>Know resources for state-specific forms (available at www.nhpco.org).</i>
	Recommend the patient identify a proxy decision-maker who should attend the next meeting.	
Have a structured discussion of scenarios with the patient	Affirm that the goal of the process is to follow the patient's wishes if the patient loses decision-making capacity.	<i>Use a structured worksheet with typical scenarios.</i>
	Elicit the patient's overall goals related to health care.	
	Elicit the patient's preferences for specific interventions in a few salient and common scenarios.	
	Help the patient define the threshold for withdrawing and withholding interventions.	<i>Begin the discussion with persistent vegetative state and consider other scenarios, such as recovery from an acute event with serious disability; then ask the patient about his or her preferences regarding specific interventions, such as ventilators, artificial nutrition, and CPR; finally, proceeding to less invasive interventions, such as blood transfusions and antibiotics.</i>
	Define the patient's preference for the role of the proxy.	
Review the patient's preferences	After the patient has made choices of interventions, review them to ensure they are consistent and the proxy is aware of them.	
Document the patient's preferences	Formally complete the advance care directive and have a witness sign it.	
	Provide a copy for the patient and the proxy.	
	Insert a copy into the patient's medical record and summarize it in a progress note.	
Update the directive	Periodically, and with major changes in health status, review the directive with the patient and make any modifications.	
Apply the directive	The directive goes into effect only when the patient becomes unable to make medical decisions for himself or herself.	
	Reread the directive to be sure about its content.	
	Discuss your proposed actions based on the directive with the proxy.	

Abbreviation: CPR, cardiopulmonary resuscitation.

will and designates a proxy is often used, and the directive should indicate clearly whether the specified patient preferences or the proxy's choice takes precedence if they conflict. Some states have begun to put into practice a "Physician Orders for Life-Sustaining Treatment (POLST)" directive, which builds on communication between providers and patients by including guidance for end-of-life care in a color-coordinated form that follows the patient across treatment settings. The procedures for completing advance care planning documents vary according to state law.

A potentially misleading distinction relates to statutory, as opposed to advisory, documents. Statutory documents are drafted to fulfill relevant state laws. Advisory documents are drafted to reflect the patient's wishes. Both are legal, the former under state law and the latter under common or constitutional law.

LEGAL ASPECTS As of 2021, 48 states and the District of Columbia had enacted living will legislation. Massachusetts and Michigan are the two states without living will legislation. Indiana has a life-prolonging procedures declaration. States differ in the requirements for advanced directives, including whether they need to be witnessed and, if so, by how many witnesses and whether they need to be notarized. Importantly, in 25 states, the laws state that the living will is not valid if a woman is pregnant. All states except Alaska have enacted durable power of attorney for health care laws that permit patients to designate a proxy decision-maker with authority to terminate life-sustaining treatments. Only in Alaska does the law prohibit proxies from terminating life-sustaining treatments for pregnant women.

The U.S. Supreme Court has ruled that patients have a constitutional right to decide any issues related to refusing or terminating medical interventions, including life-sustaining interventions, and that mentally incompetent patients can exercise this right by providing "clear and convincing evidence" of their preferences. Since advance care directives permit patients to provide such evidence, commentators agree that they are constitutionally protected. Most commentators believe that a state is required to honor any clear advance care directive, regardless of whether it is written on an "official" form. Many states have enacted laws for the explicit purpose of honoring out-of-state directives. If a patient is not using a statutory form, it may be advisable to attach a statutory form to the advance care directive being used. State-specific forms are readily available free of charge for health care providers, patients, and families through the website of the National Hospice and Palliative Care Organization (<http://www.nhpco.org>).

REIMBURSEMENT As of January 1, 2016, the Centers for Medicare and Medicaid Services amended the physician fee schedule to reimburse discussions of advance care planning under Current Procedural Terminology codes 99497 and 99498. The session must be voluntary and include an explanation of advance care planning but need not include a completed advance care document. There can be multiple bills for the discussion if it extends over several encounters. A study found that patients who engaged in a billed advance care planning encounter were more likely to be enrolled in hospice and less likely to receive intensive therapies, despite being more likely to be hospitalized in the ICU. However, a billing incentive in and of itself may not increase advance care planning discussions by clinicians. In 2016, just 1.6% of Medicare Advantage patients had a discussion of advance care planning that was billed. Factors beyond reimbursement, such as clinicians' lack of comfort and skill in carrying out advance care planning discussions and lack of time, appear to impede discussions of advance care planning.

INTERVENTIONS

■ PHYSICAL SYMPTOMS AND THEIR MANAGEMENT

Great emphasis has been placed on addressing dying patients' pain. In order to emphasize its importance, pain assessment has frequently been included as the fifth vital sign. Heightened consideration of pain has been advocated by large health care systems such as the Veterans' Administration and accrediting bodies such as The Joint Commission. Although this embrace of pain has been symbolically important,

TABLE 12-4 Common Physical and Psychological Symptoms of Terminally Ill Patients

PHYSICAL SYMPTOMS	PSYCHOLOGICAL SYMPTOMS
Pain	Anxiety
Fatigue and weakness	Depression
Dyspnea	Hopelessness
Insomnia	Meaninglessness
Dry mouth	Irritability
Anorexia	Impaired concentration
Nausea and vomiting	Confusion
Constipation	Delirium
Cough	Loss of libido
Swelling of arms or legs	
Itching	
Diarrhea	
Dysphagia	
Dizziness	
Fecal and urinary incontinence	
Numbness/tingling in hands/feet	

available data suggest that making pain the fifth vital sign does not lead to improved pain management practices. In light of the opioid crisis in the United States, the emphasis on pain management has begun to be reexamined. For instance, in 2017 draft standards, The Joint Commission recommends nonpharmacologic pain treatment as well as identification of psychosocial risk factors for addiction. Importantly, good palliative care requires much more than good pain management. The frequency of symptoms varies by disease and other factors. The most common physical and psychological symptoms among all terminally ill patients include pain, fatigue, insomnia, anorexia, dyspnea, depression, anxiety, nausea, and vomiting. In the last days of life, terminal delirium is also common. Assessments of patients with advanced cancer have shown that patients experienced an average of 11.5 different physical and psychological symptoms (**Table 12-4**).

In the vast majority of cases, evaluations to determine the etiology of these symptoms should be limited to the history and physical examination. In some cases, radiologic or other diagnostic examinations will provide sufficient benefit in directing optimal palliative care to warrant the risks, potential discomfort, and inconvenience, especially to a seriously ill patient. Only a few of the common symptoms that present difficult management issues will be addressed in this chapter. **Additional information on the management of other symptoms, such as nausea and vomiting, insomnia, and diarrhea, can be found in Chaps. 45, 31, and 46, respectively. Information on the management of patients with cancer is provided in Chap. 69.**

Pain • FREQUENCY The frequency of pain among terminally ill patients varies significantly. Cancer (~85%), congestive heart failure (CHF; ~75%), and AIDS have been associated with a higher prevalence of pain compared to other advanced illnesses, such as COPD (~45%), chronic kidney disease (~40%), and dementia (~40%). One meta-analysis of adults with advanced or terminal illness found pain prevalence of 30–94% in patients with cancer, compared to 21–77% for COPD, 14–78% for CHF, 11–83% for end-stage renal disease, 14–63% for dementia, and 30–98% for AIDS.

ETIOLOGY There are two types of pain: nociceptive and neuropathic. Nociceptive pain is further divided into somatic or visceral pain. *Somatic pain* is the result of direct mechanical or chemical stimulation of nociceptors and normal neural signaling to the brain. It tends to be localized, aching, throbbing, and cramping. The classic example is bone metastases. *Visceral pain* is caused by nociceptors in gastrointestinal (GI), respiratory, and other organ systems. It is a deep or colicky type of pain classically associated with pancreatitis, myocardial infarction, or tumor invasion of viscera. *Neuropathic pain* arises from

disordered nerve signals. It is described by patients as burning, electrical, or shock-like pain. Classic examples are post-stroke pain, tumor invasion of the brachial plexus, and herpetic neuralgia.

ASSESSMENT Pain is a subjective experience. Depending on the patient's circumstances, perspective, and physiologic condition, the same physical lesion or disease state can produce different levels of reported pain and need for pain relief. Systematic assessment includes eliciting the following: (1) type: throbbing, cramping, burning, etc.; (2) periodicity: continuous, with or without exacerbations, or incident; (3) location; (4) intensity; (5) modifying factors; (6) effects of treatments; (7) functional impact; and (8) impact on patient. Several validated pain assessment measures may be used, including the Visual Analogue Scale (VAS), the Brief Pain Inventory (BPI), or the Numerical Pain Rating Scale (NRS-11). Other scales have been developed for neuropathic pain, such as the Neuropathic Pain Scale and the DN4 Questionnaire. Frequent reassessments on a consistent scale are essential to assess the impact of and need to readjust interventions.

INTERVENTIONS Interventions for pain must be tailored to each individual, with the goal of preempting chronic pain and relieving breakthrough pain. At the end of life, there is rarely reason to doubt a patient's report of pain. With the opioid crisis in the United States, there is more emphasis on making opioids one component of multimodal analgesia. Nevertheless, at the end of life, pain medications, especially opioids, remain the cornerstone of management (Fig. 12-2). If they are failing and nonpharmacologic interventions—including radiotherapy and anesthetic or neurosurgical procedures such as peripheral nerve blocks or epidural medications—are required, a pain consultation is appropriate.

Pharmacologic interventions still largely follow the World Health Organization three-step, "analgesic ladder" approach, which involves nonopioid analgesics, "mild" opioids, and "strong" opioids, with or without adjuvants (Chap. 13). Nonopioid analgesics, especially nonsteroidal anti-inflammatory drugs (NSAIDs), are the initial treatments for mild pain. They work primarily by inhibiting peripheral prostaglandins

and reducing inflammation but may also have central nervous system (CNS) effects. Additionally, NSAIDs have a ceiling effect. Ibuprofen, up to 2400 mg/d qid, has a minimal risk of causing bleeding and renal impairment and is a good initial choice. In patients with a history of severe GI or other bleeding, however, ibuprofen should be avoided. In patients with a history of mild gastritis or gastroesophageal reflux disease (GERD), acid-lowering therapy, such as a proton pump inhibitor, should be used. Acetaminophen is an alternative in patients with a history of GI bleeding and can be used safely at up to 4 g/d qid. In patients with liver dysfunction due to metastases or other causes and in patients with heavy alcohol use, doses should be reduced.

If nonopioid analgesics are insufficient, opioids should be introduced. Opioids primarily work by interacting with μ opioid receptors to activate pain-inhibitory neurons in the CNS, although they also interact variably with δ and κ receptors. Receptor agonists, such as morphine, codeine, and fentanyl, produce analgesia by activating pain-inhibitory neurons in the CNS. Partial agonists, such as buprenorphine, have a ceiling effect for analgesia and a lower potential for abuse. They are useful for postacute pain but should not be used for chronic pain in end-of-life care. Pure antagonists, such as naloxone and methylnaltrexone, are used for reversal of opioid effects.

Traditionally, "weak" opioids such as codeine were used first. If they failed to relieve pain after dose escalation, "strong" opioids like morphine were used in doses of 5–10 mg every 4 h. However, this breakdown between "weak" and "strong" opioids is no longer commonly accepted, with smaller doses of "stronger" opioids frequently being preferred over similar or larger doses of "weaker" opioids, and different pain syndromes having different preferred therapies. Regardless, non-opioid analgesics should be combined with opioids, as they potentiate the effect of opioids.

Importantly, the goal is to prevent patients from experiencing pain. Consequently, for continuous pain, opioids should be administered on a regular, around-the-clock basis consistent with their duration of analgesia, and the next dose should occur before the effect of the previous dose wears off. They should not be provided only when the patient

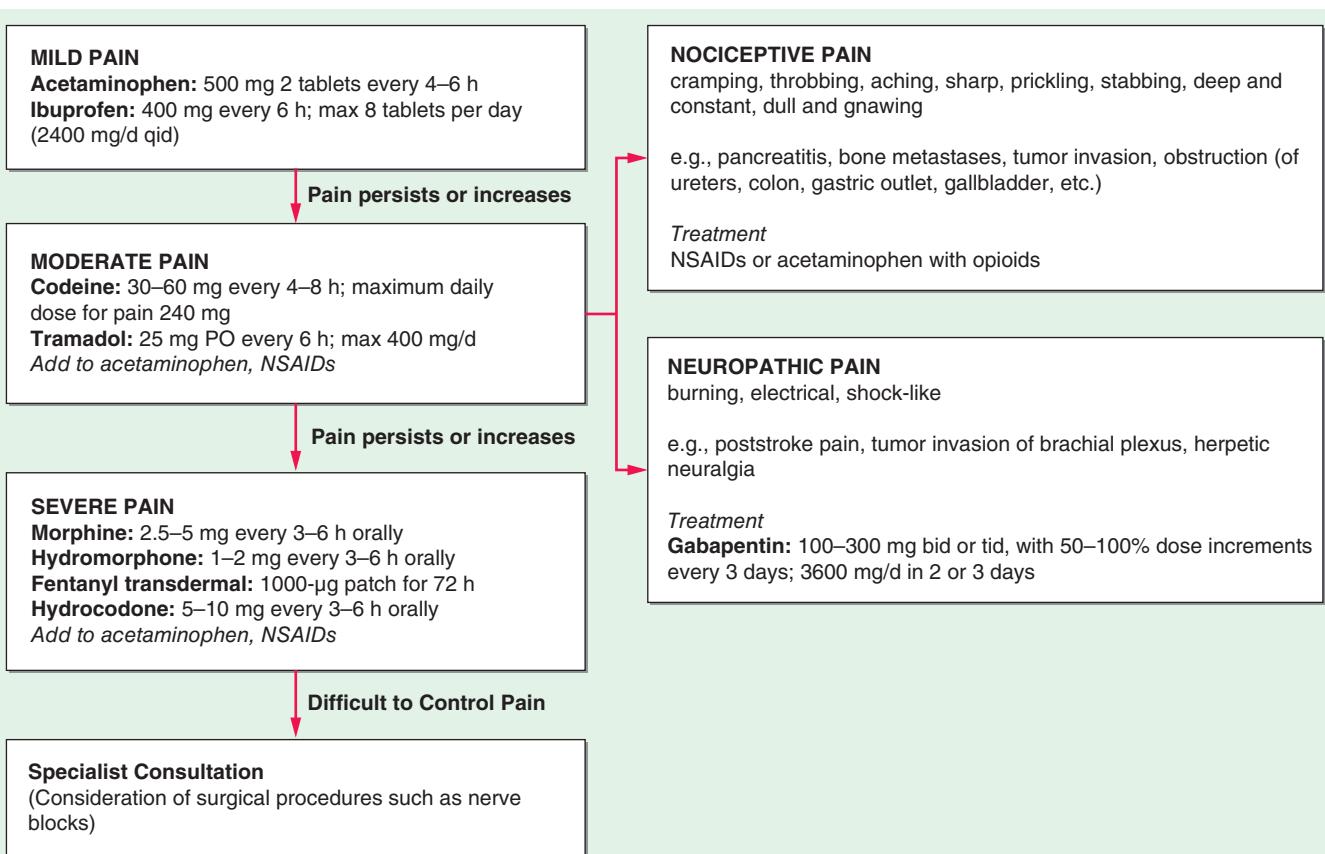


FIGURE 12-2 Terminal pain management flow chart. NSAIDs, nonsteroidal anti-inflammatory drugs.

experiences pain. Patients should also be provided rescue medication, such as liquid morphine, for breakthrough pain, generally at 20% of the baseline dose. Patients should be informed that using the rescue medication does not obviate the need to take the next standard dose of pain medication. If the patient's pain remains uncontrolled after 24 h and recurs before the next dose, requiring the patient to utilize the rescue medication, the daily opioid dose can be increased by the total dose of rescue medications used by the patient, or by 50% of the standing opioid daily dose for moderate pain and 100% for severe pain.

It is inappropriate to start with extended-release preparations. Instead, an initial focus on using short-acting preparations to determine how much is required in the first 24–48 h will allow clinicians to determine opioid needs. Once pain relief is obtained using short-acting preparations, the switch should be made to extended-release preparations. Even with a stable extended-release preparation regimen, the patient may experience incident pain, such as during movement or dressing changes. Short-acting preparations should be taken before such predictable episodes. Although less common, patients may have "end-of-dose failure" with long-acting opioids, meaning that they develop pain after 8 h in the case of an every-12-h medication. In these cases, a trial of giving an every-12-h medication every 8 h is appropriate.

Due to differences in opioid receptors, cross-tolerance among opioids is incomplete, and patients may experience different side effects with different opioids. Therefore, if a patient is not experiencing pain relief or is experiencing too many side effects, a change to another opioid preparation is appropriate. When switching, one should begin with 50–75% of the published equianalgesic dose of the new opioid.

Unlike NSAIDs, opioids have no ceiling effect; therefore, there is no maximum dose, no matter how many milligrams the patient is receiving. The appropriate dose is the dose needed to achieve pain relief. This is an important point for clinicians to explain to patients and families. Addiction or excessive respiratory depression is extremely unlikely in the terminally ill; fear of these side effects should neither prevent escalating opioid medications when the patient is experiencing insufficient pain relief nor justify using opioid antagonists.

Opioid side effects should be anticipated and treated preemptively. Nearly all patients experience constipation that can be debilitating (see below). Failure to prevent constipation often results in noncompliance with opioid therapy. The preferred treatment is prevention. Cathartics (senna 2 tbsp qHS), stool softeners (docusate 100 mg PO qd), and/or laxatives (laxulose 30 mL qd) are considered first-line treatment. For refractory cases, opioid antagonists or other therapies, such as lubiprostone, should be considered.

Methylnaltrexone is the best-studied opioid antagonist for use in refractory opioid-induced constipation. It reverses opioid-induced constipation by blocking peripheral opioid receptors, but not central receptors, for analgesia. In placebo-controlled trials, it has been shown to cause laxation within 24 h of administration. As with the use of opioids, about a third of patients using methylnaltrexone experience nausea and vomiting, but unlike with opioid usage, tolerance usually develops within a week. Therefore, when one is beginning opioids, an antiemetic such as metoclopramide or a serotonin antagonist is often prescribed prophylactically and stopped after 1 week. Olanzapine has also been shown to have antinausea properties and can be effective in countering delirium or anxiety, with the advantage of some weight gain.

Drowsiness, a common side effect of opioids, also usually abates within a week. For refractory or severe cases, pharmacologic therapy should be considered. The best-studied agents are the psychostimulants dextroamphetamine, methylphenidate, and modafinil, although evidence regarding their efficacy is weak. Modafinil has the advantage of once-a-day dosing compared to methylphenidate's twice daily dosing.

Seriously ill patients who require chronic pain relief rarely become addicted. Suspicion of addiction should not be a reason to withhold pain medications from terminally ill patients. Nonetheless, patients and families may withhold prescribed opioids for fear of addiction or dependence. Physicians and health care providers should reassure patients and families that the patient will not become addicted to

opioids if they are used as prescribed for pain relief; this fear should not prevent the patient from taking the medications around the clock. However, diversion of drugs for use by other family members or illicit sale may occur. It may be necessary to advise the patient and caregiver about secure storage of opioids. Contract writing with the patient and family can help. If that fails, transfer to a safe facility may be necessary.

Tolerance describes the need to increase medication dosage for the same pain relief without a concurrent change in disease. In the case of patients with advanced disease, the need for increasing opioid dosage for pain relief usually is caused by disease progression rather than tolerance. Physical dependence is indicated by symptoms resulting from the abrupt withdrawal of opioids and should not be confused with addiction.

In recent years, the potential dangers of opioid drugs have become increasingly apparent. To help mitigate the risk of these powerful drugs, several strategies should be used to reduce the risk of aberrant drug use. To start, all patients should be assessed for their individual levels of risk. While there are multiple surveys available, including the Opioid Risk Tool, none have gained widespread use or validation. In general, however, it is important to screen for prior substance abuse and major psychiatric disorders.

For patients deemed to be high risk, a multidisciplinary effort should be pursued to reduce the risk of adverse consequences, such as addiction and diversion. Prescribing strategies include selecting opioids with longer durations of action and lower street values, such as methadone, and prescribing smaller quantities with more frequent follow-up. Monitoring options include periodic urine screening and referral to pain specialists. In some cases, it may also be reasonable to consider not offering short-acting opioids for breakthrough pain. In no situation, however, should adequate pain relief be withheld due to risk.

Adjuvant analgesic medications are nonopioids that potentiate the analgesic effects of opioids. They are especially important in the management of neuropathic pain. Gabapentin, an anticonvulsant initially studied in the setting of herpetic neuralgia, is now the first-line treatment for neuropathic pain resulting from a variety of causes. It is begun at 100–300 mg bid or tid, with 50–100% dose increments every 3 days. Usually 900–3600 mg/d in two or three doses is effective. The combination of gabapentin and nortriptyline may be more effective than gabapentin alone. Two potential side effects of gabapentin to be aware of are confusion and drowsiness, especially in the elderly. Other effective adjuvant medications include pregabalin, which has the same mechanism of action as gabapentin but is absorbed more efficiently from the GI tract. Lamotrigine is a novel agent whose mechanism of action is unknown but has been shown to be effective. It is recommended to begin at 25–50 mg/d, increasing to 100 mg/d. Carbamazepine, a first-generation agent, has been proven effective in randomized trials for neuropathic pain. Other potentially effective anticonvulsant adjuvants include topiramate (25–50 mg qd or bid, rising to 100–300 mg/d) and oxcarbazepine (75–300 mg bid, rising to 1200 mg bid).

Glucocorticoids, preferably dexamethasone given once a day, can be useful in reducing inflammation that causes pain, while also elevating mood, energy, and appetite. Its main side effects include confusion, sleep difficulties, and fluid retention. Glucocorticoids are especially effective for bone pain and abdominal pain from distention of the GI tract or liver. Other drugs, including clonidine and baclofen, can be effective in providing pain relief. These drugs are adjuvants and generally should be used in conjunction with—not instead of—opioids. Methadone, carefully dosed because of its unpredictable half-life in many patients, has activity at the N-methyl-D-aspartate (NMDA) receptor and is useful for complex pain syndromes and neuropathic pain. It is generally reserved for cases in which first-line opioids (morphine, oxycodone, hydromorphone) are either ineffective or unavailable.

Radiation therapy can treat bone pain from single metastatic lesions. Bone pain from multiple metastases can be amenable to radiopharmaceuticals such as strontium-89 and samarium-153. Bisphosphonates, such as pamidronate (90 mg every 4 weeks) and calcitonin (200 IU intranasally once or twice a day), also provide relief from bone pain but have multiday onsets of action.

Constipation • FREQUENCY Constipation is reported in up to 70–100% of patients requiring palliative care.

ETIOLOGY Although hypercalcemia and other factors can cause constipation, it is most frequently a predictable consequence of the use of opioids for pain and dyspnea relief and of the anticholinergic effects of tricyclic antidepressants, as well as due to the inactivity and poor diets common among seriously ill patients. If left untreated, constipation can cause substantial pain and vomiting and also is associated with confusion and delirium. Whenever opioids and other medications known to cause constipation are used, preemptive treatment for constipation should be instituted.

ASSESSMENT Assessing constipation can be difficult because people describe it differently. Four commonly used assessment scales are the Bristol Stool Form Scale, the Constipation Assessment Scale, the Constipation Visual Analogue Scale, and the Eton Scale Risk Assessment for Constipation. The Bowel Function Index can be used to quantify opioid-induced constipation. The physician should establish the patient's previous bowel habits, as well as any changes in subjective and objective qualities such as bloating or decreased frequency. Abdominal and rectal examinations should be performed to exclude impaction or an acute abdomen. Radiographic assessments beyond a simple flat plate of the abdomen in cases in which obstruction is suspected are rarely necessary.

INTERVENTION Any measure to address constipation during end-of-life care should include interventions to reestablish comfortable bowel habits and to relieve pain or discomfort. Although physical activity, adequate hydration, and dietary treatments with fiber can be helpful, each is limited in its effectiveness for most seriously ill patients, and fiber may exacerbate problems in the setting of dehydration or if impaired motility is the etiology. Fiber is contraindicated in the presence of opioid use. Stimulant and osmotic laxatives, stool softeners, fluids, and enemas are the mainstays of therapy (Table 12-5). To prevent constipation from opioids and other medications, a combination of a laxative and a stool softener (such as senna and docusate) should be used. If after several days of treatment a bowel movement has not occurred, a rectal examination to remove impacted stool and place a suppository is necessary. For patients with impending bowel

obstruction or gastric stasis, octreotide to reduce secretions can be helpful. For patients in whom the suspected mechanism is dysmotility, metoclopramide can be helpful.

Nausea • FREQUENCY Up to 70% of patients with advanced cancer have nausea, defined as the subjective sensation of wanting to vomit.

ETIOLOGY Nausea and vomiting are both caused by stimulation at one of four sites: the GI tract, the vestibular system, the chemoreceptor trigger zone (CTZ), and the cerebral cortex. Medical treatments for nausea are aimed at receptors at each of these sites: the GI tract contains mechanoreceptors, chemoreceptors, and 5-hydroxytryptamine type 3 (5-HT₃) receptors; the vestibular system probably contains histamine and acetylcholine receptors; and the CTZ contains chemoreceptors, dopamine type 2 receptors, and 5-HT3 receptors. An example of nausea that most likely is mediated by the cortex is anticipatory nausea before a dose of chemotherapy or other noxious stimuli.

Specific causes of nausea include metabolic changes (liver failure, uremia from renal failure, hypercalcemia), bowel obstruction, constipation, infection, GERD, vestibular disease, brain metastases, medications (including antibiotics, NSAIDs, proton pump inhibitors, opioids, and chemotherapy), and radiation therapy. Anxiety can also contribute to nausea.

INTERVENTION Medical treatment of nausea is directed at the anatomic and receptor-mediated cause revealed by a careful history and physical examination. When no specific cause of nausea is identified, many advocate beginning treatment with metoclopramide; a serotonin type 3 (5-HT₃) receptor antagonist such as ondansetron, granisetron, palonosetron, dolasetron, tropisetron, or ramosetron; or a dopamine antagonist such as chlorpromazine, haloperidol, or prochlorperazine. When decreased motility is suspected, metoclopramide can be an effective treatment. When inflammation of the GI tract is suspected, glucocorticoids, such as dexamethasone, are an appropriate treatment. For nausea that follows chemotherapy and radiation therapy, one of the 5-HT₃ receptor antagonists or neurokinin-1 antagonists, such as aprepitant or fosaprepitant, is recommended. Clinicians should attempt prevention of postchemotherapy nausea, rather than simply providing treatment after the fact. Current clinical guidelines recommend tailoring the strength of treatments to the specific emetic risk posed by a specific chemotherapy drug. When a vestibular cause (such as "motion sickness" or labyrinthitis) is suspected, antihistamines, such as meclizine (whose primary side effect is drowsiness), or anticholinergics, such as scopolamine, can be effective. In anticipatory nausea, patients can benefit from nonpharmacologic interventions, such as biofeedback and hypnosis. The most common pharmacologic intervention for anticipatory nausea is a benzodiazepine, such as lorazepam. As with antihistamines, drowsiness and confusion are the main side effects.

The use of medical marijuana or oral cannabinoids for palliative treatment of nausea is controversial, as there are no controlled trials showing its effectiveness for patients at the end of life. A 2015 meta-analysis showed "low-quality evidence suggesting that cannabinoids were associated with improvements in nausea and vomiting due to chemotherapy," and such treatments are not as good as 5-HT₃ receptor antagonists and can sometimes even cause cannabis hyperemesis syndrome. Older patients, who compose the vast majority of dying patients, seem to tolerate cannabinoids poorly.

Dyspnea • FREQUENCY Dyspnea is the subjective experience of being short of breath. Over 50%, and as many as 75%, of dying patients, especially those with lung cancer, metastases to the lung, CHF, and COPD, experience dyspnea at some point near the end of life. Dyspnea is among the most distressing of physical symptoms and can be even more distressing than pain.

ASSESSMENT As with pain, dyspnea is a subjective experience that may not correlate with objective measures of PO₂, PCO₂, or respiratory rate. Consequently, measurements of oxygen saturation through pulse oximetry or blood gases are rarely helpful in guiding therapy. Despite the limitations of existing assessment methods, physicians should regularly assess and document patients' experience of dyspnea

TABLE 12-5 Medications for the Management of Constipation

INTERVENTION	DOSE	COMMENT
Stimulant laxatives		These agents directly stimulate peristalsis and may reduce colonic absorption of water.
Prune juice	120–240 mL/d	Work in 6–12 h.
Senna (Senokot)	2–8 tablets PO bid	
Bisacodyl	5–15 mg/d PO, PR	
Osmotic laxatives		These agents are not absorbed. They attract and retain water in the gastrointestinal tract.
Lactulose	15–30 mL PO q4–8h	Lactulose may cause flatulence and bloating.
Magnesium hydroxide (Milk of Magnesia)	15–30 mL/d PO	Lactulose works in 1 day, magnesium products in 6 h.
Magnesium citrate	125–250 mL/d PO	
Stool softeners		These agents work by increasing water secretion and as detergents, increasing water penetration into the stool.
Sodium docusate (Colace)	300–600 mg/d PO	Work in 1–3 days.
Calcium docusate	300–600 mg/d PO	
Suppositories and enemas		
Bisacodyl	10–15 PR qd	
Sodium phosphate enema	PR qd	Fixed dose, 4.5 oz, Fleet's.

and its intensity. Guidelines recommend visual analogue dyspnea scales to assess the severity of symptoms and the effects of treatment. Potentially reversible or treatable causes of dyspnea include infection, pleural effusions, pulmonary emboli, pulmonary edema, asthma, and tumor encroachment on the airway. However, the risk-versus-benefit ratio of the diagnostic and therapeutic interventions for patients with little time left to live must be considered carefully before undertaking diagnostic steps. Frequently, the specific etiology cannot be identified, and dyspnea is the consequence of progression of the underlying disease that cannot be treated. The anxiety caused by dyspnea and the choking sensation can significantly exacerbate the underlying dyspnea in a negatively reinforcing cycle.

INTERVENTIONS When reversible or treatable etiologies are diagnosed, they should be treated as long as the side effects of treatment, such as repeated drainage of effusions or anticoagulants, are less burdensome than the dyspnea itself. More aggressive treatments such as stenting a bronchial lesion may be warranted if it is clear that the dyspnea is due to tumor invasion at that site and if the patient and family understand the risks of such a procedure.

Usually, treatment will be symptomatic (**Table 12-6**). Supplemental oxygen does not appear to be effective. “A systematic review of the literature failed to demonstrate a consistent beneficial effect of oxygen inhalation over air inhalation for study participants with dyspnea due to end-stage cancer or cardiac failure.” Therefore, oxygen may be no more than an expensive placebo. Low-dose opioids reduce the sensitivity of the central respiratory center and relieve the sensation of dyspnea. If patients are not receiving opioids, weak opioids can be initiated; if patients are already receiving opioids, morphine or other stronger opioids should be used. Controlled trials do not support the use of nebulized opioids for dyspnea at the end of life. Phenothiazines and chlorpromazine may be helpful when combined with opioids. Benzodiazepines can be helpful in treating dyspnea, but only if anxiety is present. Benzodiazepines should not be used as first-line therapy or if there is no anxiety. If the patient has a history of COPD or asthma, inhaled bronchodilators and glucocorticoids may be helpful. If the patient has pulmonary edema due to heart failure, diuresis with a medication such as furosemide is indicated. Excess secretions can be transdermally or intravenously dried with scopolamine. More general interventions that medical staff can perform include sitting the patient upright, removing smoke or other irritants like perfume, ensuring a supply of fresh air with sufficient humidity, and minimizing other factors that can increase anxiety.

TABLE 12-6 Medications for the Management of Dyspnea

INTERVENTION	DOSE	COMMENTS
Weak opioids		For patients with mild dyspnea
Codeine (or codeine with 325 mg acetaminophen)	30 mg PO q4h	For opioid-naïve patients
Hydrocodone	5 mg PO q4h	
Strong opioids		For opioid-naïve patients with moderate to severe dyspnea
Morphine	5–10 mg PO q4h	For patients already taking opioids for pain or other symptoms
	30–50% of baseline opioid dose q4h	
Oxycodone	5–10 mg PO q4h	
Hydromorphone	1–2 mg PO q4h	
Anxiolytics		Give a dose every hour until the patient is relaxed; then provide a dose for maintenance
Lorazepam	0.5–2.0 mg PO/SL/IV qh then q4–6h	
Clonazepam	0.25–2.0 mg PO q12h	
Midazolam	0.5 mg IV q15min	

Fatigue • FREQUENCY Fatigue is one of the most commonly reported symptoms not only of cancer treatment but also of the palliative care of multiple sclerosis, COPD, heart failure, and HIV. More than 90% of terminally ill patients experience fatigue and/or weakness. Fatigue is frequently cited as one of the most distressing symptoms in these patients.

ETIOLOGY The multiple causes of fatigue in the terminally ill can be categorized as resulting from the underlying disease; from disease-induced factors such as tumor necrosis factor and other cytokines; and from secondary factors such as dehydration, anemia, infection, hypothyroidism, and drug side effects. In addition to low caloric intake, loss of muscle mass and changes in muscle enzymes may play an important role in fatigue during terminal illness. The importance of changes in the CNS, especially the reticular activating system, have been hypothesized based on reports of fatigue in patients receiving cranial radiation, experiencing depression, or having chronic pain in the absence of cachexia or other physiologic changes. Finally, depression and other causes of psychological distress can contribute to fatigue.

ASSESSMENT Like pain and dyspnea, fatigue is subjective, as it represents a patient’s sense of tiredness and decreased capacity for physical work. Objective changes, even in body mass, may be absent. Consequently, assessment must rely on patient self-reporting. Scales used to measure fatigue, such as the Edmonton Functional Assessment Tool, the Fatigue Self-Report Scales, and the Rhoten Fatigue Scale, are usually appropriate for research but not clinical purposes. In clinical practice, a simple performance assessment such as the Karnofsky performance status or the Eastern Cooperative Oncology Group (ECOG)’s question “How much of the day does the patient spend in bed?” may be the best measure. In the ECOG 0–4 performance status assessment, 0 = normal activity; 1 = symptomatic without being bedridden; 2 = requiring some, but <50%, bed time; 3 = bedbound more than half the day; and 4 = bedbound all the time. Such a scale allows for assessment over time and correlates with overall disease severity and prognosis. A 2008 review by the European Association of Palliative Care also described several longer assessment tools that contained 9–20 items, including the Piper Fatigue Inventory, the Multidimensional Fatigue Inventory, and the Brief Fatigue Inventory (BFI).

INTERVENTIONS Reversible causes of fatigue, such as anemia and infection, should be treated. However, at the end of life, it must be realistically acknowledged that fatigue will not be “cured.” The goal is to ameliorate fatigue and help patients and families adjust expectations. Behavioral interventions should be utilized to avoid blaming the patient for inactivity and to educate both the family and the patient that the underlying disease causes physiologic changes that produce low energy levels. Understanding that the problem is physiologic and not psychological can help alter expectations regarding the patient’s level of physical activity. Practically, this may mean reducing routine activities such as housework, cooking, and social events outside the house and making it acceptable to receive guests while lying on a couch. At the same time, the implementation of exercise regimens and physical therapy can raise endorphins, reduce muscle wasting, and decrease the risk of depression. In addition, ensuring good hydration without worsening edema may help reduce fatigue. Discontinuing medications that worsen fatigue may help, including cardiac medications, benzodiazepines, certain antidepressants, or opioids if the patient’s pain is well-controlled. As end-of-life care proceeds into its final stages, fatigue may protect patients from further suffering, and continued treatment could be detrimental.

Only a few pharmacologic interventions target fatigue and weakness. Randomized controlled trials suggest glucocorticoids can increase energy and enhance mood. Dexamethasone (8 mg/d) is preferred for its once-a-day dosing and minimal mineralocorticoid activity. Benefit, if any, is usually seen within the first month. For fatigue related to anorexia, megestrol (480–800 mg) can be helpful. Psychostimulants such as dextroamphetamine (5–10 mg PO) and methylphenidate (2.5–5 mg PO) may enhance energy levels, although controlled trials have not shown these drugs to be effective for fatigue induced by mild

to moderate cancer. Doses should be given in the morning and at noon to minimize the risk of counterproductive insomnia. Modafinil and armodafinil, developed for narcolepsy, have shown promise in the treatment of fatigue and have the advantage of once-daily dosing. Their precise role in fatigue at the end of life has not been documented but may be worth trying if other interventions are not beneficial. Anecdotal evidence suggests that L-carnitine may improve fatigue, depression, and sleep disruption.

PALLIATIVE SEDATION

Palliative sedation is used in distressing situations that cannot be addressed in other ways. When patients experience severe symptoms, such as pain or dyspnea, that cannot be relieved by conventional interventions or experience acute catastrophic symptoms, such as uncontrolled seizures, then palliative sedation should be considered as an intervention of last resort. It can be abused if done to hasten death (which it usually does not), when done at the request of the family rather than according to the patient's wishes, or when there are other interventions that could still be tried. The use of palliative sedation in cases of extreme existential or spiritual distress remains controversial. Typically, palliative sedation should be introduced only after the patient and family have been assured that all other interventions have been tried and after the patient and their loved ones have been able to "say goodbye."

Palliative sedation can be achieved by significantly increasing opioid doses until patients become unconscious and then putting them on a continuous infusion. Another commonly used medication for palliative sedation is midazolam at 1–5 mg IV every 5–15 min to calm the patient, followed by a continuous IV or subcutaneous infusion of 1 mg/h. In hospital settings, a continuous propofol infusion of 5 µg/kg per min can be used. There are also other, less commonly used medications for palliative sedation that include levomepromazine, chlorpromazine, and phenobarbital.

PSYCHOLOGICAL SYMPTOMS AND THEIR MANAGEMENT

Depression • FREQUENCY AND IMPACT Depression at the end of life presents an apparently paradoxical situation. Many people believe that depression is normal among seriously ill patients because they are dying. People frequently say, "Wouldn't you be depressed?" Although sadness, anxiety, anger, and irritability are normal responses to a serious condition, they are typically of modest intensity and transient. Persistent sadness and anxiety and the physically disabling symptoms that they can lead to are abnormal and suggestive of major depression. The precise number of terminally ill patients who are depressed is uncertain, primarily due to a lack of consistent diagnostic criteria and screening. Careful follow-up of patients suggests that while as many as 75% of terminally ill patients experience depressive symptoms, ~25% of terminally ill patients have major depression. Depression at the end of life is concerning because it can decrease the quality of life, interfere with closure in relationships and other separation work, obstruct adherence to medical interventions, and amplify the suffering associated with pain and other symptoms.

ETIOLOGY Previous history of depression, family history of depression or bipolar disorder, and prior suicide attempts are associated with increased risk for depression among terminally ill patients. Other symptoms, such as pain and fatigue, are associated with higher rates of depression; uncontrolled pain can exacerbate depression, and depression can cause patients to be more distressed by pain. Many medications used in the terminal stages, including glucocorticoids, and some anticancer agents, such as tamoxifen, interleukin 2, interferon α, and vincristine, also are associated with depression. Some terminal conditions, such as pancreatic cancer, certain strokes, and heart failure, have been reported to be associated with higher rates of depression, although this is controversial. Finally, depression may be attributable to grief over the loss of a role or function, social isolation, or loneliness.

ASSESSMENT Unfortunately, many studies suggest that most depressed patients at the end of life are not diagnosed, or if they are diagnosed,

they are not properly treated. Diagnosing depression among seriously ill patients is complicated, as many of the vegetative symptoms in the *DSM-V (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition)* criteria for clinical depression—insomnia, anorexia and weight loss, fatigue, decreased libido, and difficulty concentrating—are associated with the process of dying itself. The assessment of depression in seriously ill patients therefore should focus on the dysphoric mood, helplessness, hopelessness, and lack of interest, enjoyment, and concentration in normal activities. It is now recommended that patients near the end of life should be screened with either the PHQ-9 or the PHQ-2, which asks "Over the past 2 weeks, how often have you been bothered by any of the following problems? (1) Little interest or pleasure in doing things and (2) feeling down, depressed or hopeless." The answer categories are as follows: not at all, several days, more than half the days, nearly every day. Other possible diagnostic tools include the short form of the Beck Depression Index or a visual analogue scale.

Certain conditions may be confused with depression. Endocrinopathies, such as hypothyroidism and Cushing's syndrome, electrolyte abnormalities, such as hypercalcemia, and akathisia, especially from dopamine-blocking antiemetics such as metoclopramide and prochlorperazine, can mimic depression and should be excluded.

INTERVENTIONS Undertreatment of depressed, terminally ill patients is common. Physicians must treat any physical symptom, such as pain, that may be causing or exacerbating depression. Fostering adaptation to the many losses that the patient is experiencing can also be helpful. Unfortunately, there are few randomized trials to guide such interventions. Thus, treatment typically follows the treatment used for non-terminally ill depressed patients.

In the absence of randomized controlled trials, nonpharmacologic interventions, including group or individual psychological counseling, and behavioral therapies such as relaxation and imagery can be helpful, especially in combination with drug therapy.

Pharmacologic interventions remain at the core of therapy. The same medications are used to treat depression in terminally ill as in non-terminally ill patients. Psychostimulants may be preferred for patients with a poor prognosis or for those with fatigue or opioid-induced somnolence. Psychostimulants are comparatively fast-acting, working within a few days instead of the weeks required for selective serotonin reuptake inhibitors (SSRIs). Dextroamphetamine or methylphenidate should be started at 2.5–5.0 mg in the morning and at noon, the same starting doses used for treating fatigue. The doses can eventually be escalated up to 15 mg bid. Modafinil is started at 100 mg qd and can be increased to 200 mg if there is no effect at the lower dose. Pemoline is a nonamphetamine psychostimulant with minimal abuse potential. It is also effective as an antidepressant beginning at 18.75 mg in the morning and at noon. Because it can be absorbed through the buccal mucosa, it is preferred for patients with intestinal obstruction or dysphagia. If it is used for prolonged periods, liver function must be monitored. The psychostimulants can also be combined with more traditional antidepressants while waiting for the antidepressants to become effective, then tapered down after a few weeks if necessary. Psychostimulants have side effects, particularly initial anxiety, insomnia, and very rarely paranoia, which may necessitate lowering the dose or discontinuing treatment.

Mirtazapine, an antagonist at the postsynaptic serotonin receptors, is a promising psychostimulant. It should be started at 7.5 mg before bed and titrated up no more than once every 1–2 weeks to a maximal dose of 45 mg/d. It has sedating, antiemetic, and anxiolytic properties, with few drug interactions. Its side effect of weight gain may be beneficial for seriously ill patients; it is available in orally disintegrating tablets.

For patients with a prognosis of several months or longer, SSRIs, including fluoxetine, sertraline, paroxetine, escitalopram, and citalopram, and serotonin-noradrenaline reuptake inhibitors, such as venlafaxine and duloxetine, are the preferred treatments, due to their efficacy and comparatively few side effects. Because low doses of these medications may be effective for seriously ill patients, one should use half the usual starting dose as for healthy adults. The starting dose for

fluoxetine is 10 mg once a day. In most cases, once-a-day dosing is possible. The choice of which SSRI to use should be driven by (1) the patient's past success or failure with the specific medication and (2) the most favorable side effect profile for that specific agent. For instance, for a patient in whom fatigue is a major symptom, a more activating SSRI (fluoxetine) would be appropriate. For a patient in whom anxiety and sleeplessness are major symptoms, a more sedating SSRI (paroxetine) would be appropriate. Importantly, it can take up to 4 weeks for these drugs to have an effect.

Atypical antidepressants are recommended only in select circumstances, usually with the assistance of a specialty consultation. Trazodone can be an effective antidepressant but is sedating and can cause orthostatic hypotension and, occasionally, priapism. Therefore, it should be used before bed and only when a sedating effect is desired and is often used for patients with insomnia at a dose starting at 25 mg. Bupropion can also be used. In addition to its antidepressant effects, bupropion is energizing, making it useful for depressed patients who experience fatigue. However, it can cause seizures, preventing its use for patients with a risk of CNS neoplasms or terminal delirium. Finally, alprazolam, a benzodiazepine, starting at 0.25–1.0 mg tid, can be effective in seriously ill patients who have a combination of anxiety and depression. Although it is potent and works quickly, it has many drug interactions and may cause delirium, especially among very ill patients, because of its strong binding to the benzodiazepine- γ -aminobutyric acid (GABA) receptor complex.

Unless used as adjuvants for the treatment of pain, tricyclic antidepressants are not recommended. While they can be effective, their therapeutic window and serious side effects typically limit their utility. Similarly, monoamine oxidase (MAO) inhibitors are not recommended because of their side effects and dangerous drug interactions.

Delirium (See Chap. 27) • **FREQUENCY** In the weeks or months before death, delirium is uncommon, although it may be significantly underdiagnosed. However, delirium becomes relatively common in the days and hours immediately before death. Up to 85% of patients dying from cancer may experience terminal delirium.

ETIOLOGY Delirium is a global cerebral dysfunction characterized by alterations in cognition and consciousness. It is frequently preceded by anxiety, changes in sleep patterns (especially reversal of day and night), and decreased attention. In contrast to dementia, delirium has an acute onset, is characterized by fluctuating consciousness and inattention, and is reversible, although reversibility may be more theoretical than real for patients near death. Delirium may occur in a patient with dementia; indeed, patients with dementia are more vulnerable to delirium.

Causes of delirium include metabolic encephalopathy arising from liver or renal failure, hypoxemia, or infection; electrolyte imbalances such as hypercalcemia; paraneoplastic syndromes; dehydration; and primary brain tumors, brain metastases, or leptomeningeal spread of tumor. Among dying patients, delirium is commonly caused by side effects of treatments, including radiation for brain metastases and medications, such as opioids, glucocorticoids, anticholinergic drugs, antihistamines, antiemetics, benzodiazepines, and chemotherapeutic agents. The etiology may be multifactorial; e.g., dehydration may exacerbate opioid-induced delirium.

ASSESSMENT Delirium should be recognized in any terminally ill patient exhibiting new onset of disorientation, impaired cognition, somnolence, fluctuating levels of consciousness, or delusions with or without agitation. Delirium must be distinguished from acute anxiety, depression, and dementia. The central distinguishing feature is altered consciousness, which usually is not noted in anxiety, depression, or dementia. Although "hyperactive" delirium, characterized by overt confusion and agitation, is probably more common, patients should also be assessed for "hypoactive" delirium, which is characterized by sleep-wake reversal and decreased alertness.

In some cases, use of formal assessment tools such as the Mini-Mental Status Examination (which does not distinguish delirium from dementia) and the Delirium Rating Scale (which does distinguish

delirium from dementia) may be helpful in distinguishing delirium from other processes. The patient's list of medications must be evaluated carefully. Nonetheless, a reversible etiologic factor for delirium is found in fewer than half of all terminally ill patients. Given that most terminally ill patients experiencing delirium are very close to death and often at home, extensive diagnostic evaluations such as lumbar punctures and neuroradiologic examinations are inappropriate.

INTERVENTIONS One of the most important objectives of terminal care is to provide terminally ill patients the lucidity to say goodbye to the people they love. Delirium, especially when in combination with agitation during the final days, is distressing to family and caregivers. A strong determinant of bereavement difficulties is witnessing a difficult death. Thus, terminal delirium should be treated aggressively.

At the first sign of delirium, such as day-night reversal with slight changes in mentation, the physician should let the family members know that it is time to be sure that everything they want to say has been said. The family should be informed that delirium is common just before death.

If medications are suspected of being a cause of the delirium, unnecessary agents should be discontinued. Other potentially reversible causes, such as constipation, urinary retention, and metabolic abnormalities, should be treated. Supportive measures aimed at providing a familiar environment should be instituted, including restricting visits only to individuals with whom the patient is familiar and eliminating new experiences; orienting the patient, if possible, by providing a clock and calendar; and gently correcting the patient's hallucinations or cognitive mistakes.

Pharmacologic management focuses on the use of neuroleptics and, in extreme cases, anesthetics (Table 12-7). Haloperidol remains the first-line therapy. Usually, patients can be controlled with a low dose (1–3 mg/d), given every 6 h, although some may require as much as 20 mg/d. Haloperidol can be administered PO, SC, or IV. IM injections should not be used, except when this is the only way to address a patient's delirium. Olanzapine, an atypical neuroleptic, has shown significant effectiveness in completely resolving delirium in cancer patients. It also has other beneficial effects for terminally ill patients, including antinausea, antianxiety, and weight gain. Olanzapine is useful for patients with longer anticipated life expectancies because it is less likely to cause dysphoria and has a lower risk of dystonic reactions. Additionally, because olanzapine is metabolized through multiple pathways, it can be used in patients with hepatic and renal dysfunction. Olanzapine has the disadvantage that it is only available orally and takes a week to reach steady state. The usual dose is 2.5–5 mg PO bid. Chlorpromazine (10–25 mg every 4–6 h) can be useful if sedation is desired and can be administered IV or PR in addition to PO. Dystonic reactions resulting from dopamine blockade are a side effect of neuroleptics, although they are reported to be rare when these drugs are used to treat terminal delirium. If patients develop dystonic reactions, benztropine should be administered. Neuroleptics may be

TABLE 12-7 Medications for the Management of Delirium

INTERVENTIONS	DOSE
Neuroleptics	
Haloperidol	0.5–5 mg q2–12h, PO/IV/SC/IM
Thioridazine	10–75 mg q4–8h, PO
Chlorpromazine	12.5–50 mg q4–12h, PO/IV/IM
Atypical neuroleptics	
Olanzapine	2.5–5 mg qd or bid, PO
Risperidone	1–3 mg q12h, PO
Anxiolytics	
Lorazepam	0.5–2 mg q1–4h, PO/IV/IM
Midazolam	1–5 mg/h continuous infusion, IV/SC
Anesthetics	
Propofol	0.3–2.0 mg/h continuous infusion, IV

combined with lorazepam to reduce agitation when the delirium is the result of alcohol or sedative withdrawal.

If no response to first-line therapy is observed, a specialty consultation should be obtained with a goal to change to a different medication. If the patient fails to improve after a second neuroleptic, sedation with either an anesthetic such as propofol or continuous-infusion midazolam may be necessary. By some estimates, as many as 25% of patients at the very end of life who experience delirium, especially restless delirium with myoclonus or convulsions, may require sedation.

Physical restraints should be used with great reluctance and only when patients' violence is threatening to themselves or others. If restraints are used, their appropriateness should be frequently reevaluated.

Insomnia • FREQUENCY Sleep disorders, defined as difficulty initiating sleep or maintaining sleep, sleep difficulty at least 3 nights a week, or sleep difficulty that causes impairment of daytime functioning, occurs in 19–63% of patients with advanced cancer. Some 30–74% of patients with other end-stage conditions, including AIDS, heart disease, COPD, and renal disease, experience insomnia.

Etiology Patients with cancer may experience changes in sleep efficiency, such as an increase in stage I sleep. Insomnia may also coexist with both physical illnesses, like thyroid disease, and psychological illnesses, like depression and anxiety. Medications, including antidepressants, psychostimulants, glucocorticoids, and β agonists, are significant contributors to sleep disorders, as are caffeine and alcohol. Multiple over-the-counter medications contain caffeine and antihistamines, which can contribute to sleep disorders.

Assessment Assessments should include specific questions concerning sleep onset, sleep maintenance, and early-morning wakening, as these will provide clues to both the causative agents and management of insomnia. Patients should be asked about previous sleep problems, screened for depression and anxiety, and asked about symptoms of thyroid disease. Caffeine and alcohol are prominent causes of sleep problems, and a careful history of the use of these substances should be obtained. Both excessive use and withdrawal from alcohol can be causes of sleep problems.

Interventions The mainstays of any intervention include improvement of sleep hygiene (encouragement of regular time for sleep, decreased nighttime distractions, elimination of caffeine and other stimulants and alcohol), interventions to treat anxiety and depression, and treatment for the insomnia itself. For patients with depression who have insomnia and anxiety, a sedating antidepressant such as mirtazapine can be helpful. In the elderly, trazodone, beginning at 25 mg at nighttime, is an effective sleep aid at doses lower than those that cause its antidepressant effect. Zolpidem may have a decreased incidence of delirium in patients compared with traditional benzodiazepines, but this has not been clearly established. When benzodiazepines are prescribed, short-acting ones (such as lorazepam) are favored over longer-acting ones (such as diazepam). Patients who receive these medications should be observed for signs of increased confusion and delirium.

SOCIAL NEEDS AND THEIR MANAGEMENT

Financial Burdens • FREQUENCY Dying can impose substantial economic strains on patients and families, potentially causing distress. This is known as financial toxicity. In the United States, which has the least comprehensive health insurance systems among wealthy countries, a quarter of families coping with end-stage cancer report that care was a major financial burden and a third used up most of their savings. Among Medicare beneficiaries, average out-of-pocket costs were >\$8000. Between 10% and 30% of families are forced to sell assets, use savings, or take out a mortgage to pay for the patient's health care costs.

The patient is likely to reduce hours worked and eventually stop working altogether. In 20% of cases, a family member of the terminally ill patient also must stop working to provide care. The major underlying causes of economic burden are related to poor physical functioning and care needs, such as the need for housekeeping, nursing, and personal care. More debilitated patients and poor patients experience greater economic burdens.

Intervention The economic burden of end-of-life care should not be ignored as a private matter. It has been associated with a number of adverse health outcomes, including preferring comfort care over life-prolonging care, as well as consideration of euthanasia or physician-assisted suicide (PAS). Economic burdens increase the psychological distress of the families and caregivers of terminally ill patients, and poverty is associated with many adverse health outcomes. Importantly, studies have found that "patients with advanced cancer who reported having end-of-life conversations with physicians had significantly lower health care costs in their final week of life. Higher costs were associated with worse quality of death." Assistance from a social worker, early on if possible, to ensure access to all available benefits may be helpful. Many patients, families, and health care providers are unaware of options for long-term care insurance, respite care, the Family Medical Leave Act (FMLA), and other sources of assistance. Some of these options (such as respite care) may be part of a formal hospice program, but others (such as the FMLA) do not require enrollment in a hospice program.

Relationships • FREQUENCY Settling personal issues and closing the narrative of lived relationships are universal needs. When asked if sudden death or death after an illness is preferable, respondents often initially select the former, but soon change to the latter as they reflect on the importance of saying goodbye. Bereaved family members who have not had the chance to say goodbye often have a more difficult grief process.

Interventions Care of seriously ill patients requires efforts to facilitate the types of encounters and time spent with family and friends that are necessary to meet those needs. Family and close friends may need to be accommodated in hospitals and other facilities with unrestricted visiting hours, which may include sleeping near the patient, even in otherwise regimented institutional settings. Physicians and other health care providers may be able to facilitate and resolve strained interactions between the patient and other family members. Assistance for patients and family members who are unsure about how to create or help preserve memories, whether by providing materials such as a scrapbook or memory box or by offering them suggestions and informational resources, can be deeply appreciated. Taking photographs and creating videos can be especially helpful to terminally ill patients who have younger children or grandchildren.

Family Caregivers • FREQUENCY Caring for seriously ill patients places a heavy burden on families. Families are frequently required to provide transportation and homemaking, as well as other services. Typically, paid professionals, such as home health nurses and hospice workers, supplement family care; only about a quarter of all caregiving consists of exclusively paid professional assistance. Over the past 40 years, there has been a significant decline in the United States of deaths occurring in hospitals, with a simultaneous increase in deaths in other facilities and at home. Over a third of deaths occur in patients' homes. This increase in out-of-hospital deaths increases reliance on families for end-of-life care. Increasingly, family members are being called upon to provide physical care (such as moving and bathing patients) and medical care (such as assessing symptoms and giving medications) in addition to emotional care and support.

Three-quarters of family caregivers of terminally ill patients are women—wives, daughters, sisters, and even daughters-in-law. Since many are widowed, women tend to be able to rely less on family for caregiving assistance and may need more paid assistance. About 20% of terminally ill patients report substantial unmet needs for nursing and personal care. The impact of caregiving on family caregivers is substantial: both bereaved and current caregivers have a higher mortality rate than that of non-caregiving controls.

Interventions It is imperative to inquire about unmet needs and to try to ensure that those needs are met either through the family or by paid professional services when possible. Community assistance through houses of worship or other community groups often can be mobilized by telephone calls from the medical team to someone the patient or family identifies. Sources of support specifically for family

caregivers should be identified through local sources or nationally through groups such as the National Family Caregivers Association (www.nfcacares.org), the American Cancer Society (www.cancer.org), and the Alzheimer's Association (www.alz.org).

■ EXISTENTIAL NEEDS AND THEIR MANAGEMENT

Frequency Religion and spirituality are often important to dying patients. Nearly 70% of patients report becoming more religious or spiritual when they became terminally ill, and many find comfort in religious or spiritual practices such as prayer. However, ~20% of terminally ill patients become less religious, frequently feeling cheated or betrayed by becoming terminally ill. For other patients, the need is for existential meaning and purpose that is distinct from, and may even be antithetical to, religion or spirituality. When asked, patients and family caregivers frequently report wanting their professional caregivers to be more attentive to religion and spirituality.

Assessment Health care providers are often hesitant about involving themselves in the religious, spiritual, and existential experiences of their patients because it may seem private or not relevant to the current illness. But physicians and other members of the care team should be able at least to detect spiritual and existential needs. Screening questions have been developed for a physician's spiritual history taking. Spiritual distress can amplify other types of suffering and even masquerade as intractable physical pain, anxiety, or depression. The screening questions in the comprehensive assessment are usually sufficient. Deeper evaluation and intervention are rarely appropriate for the physician unless no other member of a care team is available or suitable. Pastoral care providers may be helpful, whether from the medical institution or from the patient's own community.

Interventions Precisely how religious practices, spirituality, and existential explorations can be facilitated and improve end-of-life care is not well established. What is clear is that for physicians, one main intervention is to inquire about the role and importance of spirituality and religion in a patient's life. This will help a patient feel heard and help physicians identify specific needs. In one study, only 36% of respondents indicated that a clergy member would be comforting. Nevertheless, the increase in religious and spiritual interest among a substantial fraction of dying patients suggests inquiring of individual patients how this need can be addressed. Some evidence supports specific methods of addressing existential needs in patients, ranging from establishing a supportive group environment for terminal patients to individual treatments emphasizing a patient's dignity and sources of meaning.

MANAGING THE LAST STAGES

■ PALLIATIVE CARE SERVICES: HOW AND WHERE

Determining the best approach to providing palliative care to patients will depend on patient preferences, the availability of caregivers and specialized services in close proximity, institutional resources, and reimbursement. Hospice is a leading, but not the only, model of palliative care services. In the United States, slightly more than a third—35.7%—of hospice care is provided in private residential homes with 14.5% of hospice care in nursing homes. In the United States, Medicare pays for hospice services under Part A, the hospital insurance part of reimbursement. Two physicians must certify that the patient has a prognosis of ≤6 months if the disease runs its usual course. Prognoses are probabilistic by their nature; patients are not required to die within 6 months but rather to have a condition from which half the individuals with it would not be alive within 6 months. Patients sign a hospice enrollment form that states their intent to forgo curative services related to their terminal illness but can still receive medical services for other comorbid conditions. Patients also can withdraw enrollment and reenroll later; the hospice Medicare benefit can be revoked later to secure traditional Medicare benefits. Payments to the hospice are per diem (or capitated), not fee-for-service. Payments are intended to cover physician services for the medical direction of the care team;

regular home care visits by registered nurses and licensed practical nurses; home health aide and homemaker services; chaplain services; social work services; bereavement counseling; and medical equipment, supplies, and medications. No specific therapy is excluded, and the goal is for each therapy to be considered for its symptomatic (as opposed to disease-modifying) effect. Additional clinical care, including services of the primary physician, is covered by Medicare Part B even while the hospice Medicare benefit is in place.

The Affordable Care Act directs the secretary of Health and Human Services to gather data on Medicare hospice reimbursement with the goal of reforming payment rates to account for resource use over an entire episode of care. The legislation also requires additional evaluations and reviews of eligibility for hospice care by hospice physicians or nurses. The Center for Medicare and Medicaid Innovation (CMMI) sponsors and carries out demonstration projects to test models and evaluate the potential of new methods. In 2016, CMMI started a 5-year test of concurrent hospice and palliative care services with curative treatment for terminally ill patients who have a life expectancy of ≤6 months. A 4-year test initiated in 2021 will examine the inclusion of hospice in Medicare Advantage covering 8% of the market and include important health plans.

By 2018, the average length of enrollment in a hospice for Medicare beneficiaries was 90 days. However, the median length of stay was just 18 days, suggesting most patients are in hospice for a short time. Such short stays create barriers to establishing high-quality palliative services in patients' homes and also place financial strains on hospice providers since the initial assessments are resource intensive. Physicians should initiate early referrals to the hospice to allow more time for patients to receive palliative care.

In the United States, hospice care has been the main method for securing palliative services for terminally ill patients. However, leading physicians have increasingly emphasized the need to introduce palliative care much earlier in patients' illness, and efforts are being made to develop palliative care services that can be provided before the last 6 months of life and across a variety of settings. Studies of terminally ill patients indicate that those who received in-home palliative care delivered by an interdisciplinary team compared to usual care were more satisfied, more likely to die at home, and had fewer visits to the emergency room and lower per-day costs. More companies and home health agencies are now offering nonhospice palliative care services in patients' homes in an effort to increase quality of life and forestall emergency room visits and hospitalizations. Similarly, palliative care services are increasingly available via consultation, rather than being available only in hospital, day care, outpatient, and nursing home settings. Palliative care consultations for nonhospice patients can be billed as for other consultations under Medicare Part B. It is argued that using palliative care earlier in patients' illness allows patients and family members to become more acculturated to avoiding life-sustaining treatments, facilitating a smoother transition to hospice care closer to death.

■ WITHDRAWING AND WITHHOLDING LIFE-SUSTAINING TREATMENT

Legal Aspects For centuries, it has been deemed ethical to withhold or withdraw life-sustaining interventions. The current legal consensus in the United States and most wealthy countries is that patients have a moral as well as legal right to refuse medical interventions. American courts also have held that incompetent patients have a right to refuse medical interventions. For patients who are incompetent and terminally ill and who have not completed an advance care directive, next of kin can exercise that right, although this may be restricted in some states, depending on how clear and convincing the evidence is of the patient's preferences. Courts have limited families' ability to terminate life-sustaining treatments in patients who are conscious and incompetent but not terminally ill. In theory, patients' right to refuse medical therapy can be limited by four countervailing interests: (1) preservation of life, (2) prevention of suicide, (3) protection of third parties such as children, and (4) preservation of the integrity of the

medical profession. In practice, these interests almost never override the right of competent patients and incompetent patients who have left explicit wishes or advance care directives.

For incompetent patients who either appointed a proxy without specific indications of their wishes or never completed an advance care directive, three criteria have been suggested to guide the decision to terminate medical interventions. First, some commentators suggest that ordinary care should be administered but extraordinary care could be terminated. Because the ordinary/extraordinary distinction is too vague, courts and commentators widely agree that it should not be used to justify decisions about stopping treatment. Second, many courts have advocated the use of the substituted-judgment criterion, which holds that the proxy decision-makers should try to imagine what the incompetent patient would do if he or she were competent. However, multiple studies indicate that many proxies, even close family members, cannot accurately predict what the patient would have wanted. Therefore, substituted judgment becomes more of a guessing game than a way of fulfilling the patient's wishes. Finally, the best-interests criterion holds that proxies should evaluate treatments by balancing their benefits and risks and select those treatments where the benefits maximally outweigh the burdens of treatment. Clinicians have a clear and crucial role in this by carefully and dispassionately explaining the known benefits and burdens of specific treatments. Yet even when that information is as clear as possible, different individuals can have very different views of what is in the patient's best interests, and families may have disagreements or even overt conflicts. This criterion has been criticized because there is no single way to determine the balance between benefits and burdens; it depends on a patient's personal values. For instance, for some people, being alive even if mentally incapacitated is a benefit, whereas for others, it may be the worst possible existence. As a matter of practice, physicians rely on family members to make decisions that they feel are best and object only if those decisions seem to demand treatments that the physicians consider not beneficial.

Practices Withholding and withdrawing acutely life-sustaining medical interventions from terminally ill patients are now standard practice. More than 90% of American patients die without cardiopulmonary resuscitation (CPR), and just as many forgo other potentially life-sustaining interventions. For instance, in ICUs in the period of 1987–1988, CPR was performed 49% of the time, but it was performed only 10% of the time in 1992–1993 and on just 1.8% of admissions from 2001 to 2008. On average, 3.8 interventions, such as vasopressors and transfusions, were stopped for each dying ICU patient. However, up to 19% of decedents in hospitals received interventions such as extubation, ventilation, and surgery in the 48 h preceding death. There is wide variation in practices among hospitals and ICUs, suggesting an important element of physician preferences rather than consistent adherence to professional society recommendations.

Mechanical ventilation may be the most challenging intervention to withdraw. The two approaches are *terminal extubation*, which is the removal of the endotracheal tube, and *terminal weaning*, which is the gradual reduction of the fraction of inspired oxygen (FIO₂) or ventilator rate. One-third of ICU physicians prefer to use the terminal weaning technique, and 13% extubate; the majority of physicians utilize both techniques. The American Thoracic Society's 2008 clinical policy guidelines note that there is no single correct process of ventilator withdrawal and that physicians use and should be proficient in both methods but that the chosen approach should carefully balance benefits and burdens as well as patient and caregiver preferences. Some recommend terminal weaning because patients do not develop upper airway obstruction and the distress caused by secretions or stridor; however, terminal weaning can prolong the dying process and not allow a patient's family to be with the patient unencumbered by an endotracheal tube. To ensure comfort for conscious or semiconscious patients before withdrawal of the ventilator, neuromuscular blocking agents should be terminated and sedatives and analgesics administered. Removing the neuromuscular blocking agents permits patients to show discomfort, facilitating the titration of sedatives and analgesics; it also permits interactions between patients and their families. A common

practice is to inject a bolus of midazolam (2–4 mg) or lorazepam (2–4 mg) before withdrawal, followed by a bolus of 5–10 mg of morphine and continuous infusion of morphine (50% of the bolus dose per hour) during weaning. In patients who have significant upper airway secretions, IV scopolamine at a rate of 100 µg/h can be administered. Additional boluses of morphine or increases in the infusion rate should be administered for respiratory distress or signs of pain. Higher doses will be needed for patients already receiving sedatives and opioids.

The median time to death after stopping of the ventilator is 1 h. However, up to 10% of patients unexpectedly survive for 1 day or more after mechanical ventilation is stopped. Women and older patients tend to survive longer after extubation. Families need to be reassured about both the continuations of treatments for common symptoms, such as dyspnea and agitation, after withdrawal of ventilatory support and the uncertainty of length of survival after withdrawal of ventilatory support.

FUTILE CARE

Beginning in the late 1980s, some commentators argued that physicians could terminate futile treatments demanded by the families of terminally ill patients. Although no objective definition or standard of futility exists, several categories have been proposed. Physiologic futility means that an intervention will have no physiologic effect. Some have defined qualitative futility as applying to procedures that "fail to end a patient's total dependence on intensive medical care." Quantitative futility occurs "when physicians conclude (through personal experience, experiences shared with colleagues, or consideration of reported empiric data) that in the last 100 cases, a medical treatment has been useless." The term conceals subjective value judgments about when a treatment is "not beneficial." Deciding whether a treatment that obtains an additional 6 weeks of life or a 1% survival advantage confers benefit depends on patients' preferences and goals. Furthermore, physicians' predictions of when treatments are futile deviate markedly from the quantitative definition. When residents thought CPR was quantitatively futile, more than one in five patients had a >10% chance of survival to hospital discharge. Most studies that purport to guide determinations of futility are based on insufficient data and therefore cannot provide statistical confidence for clinical decision-making. Quantitative futility rarely applies in ICU settings.

Many commentators reject using futility as a criterion for withdrawing care, preferring instead to consider futility situations as ones that represent conflict that calls for careful negotiation between families and health care providers. The American Medical Association and other professional societies have developed process-based approaches to resolving cases clinicians feel are futile. These process-based measures mainly suggest involving consultants and/or ethics committees when there are seemingly irresolvable differences. Some hospitals have enacted "unilateral do-not-resuscitate" policies to allow clinicians to provide a do-not-resuscitate order in cases in which consensus cannot be reached with families and medical opinion is that resuscitation would be futile if attempted. This type of a policy is not a replacement for careful and patient communication and negotiation but recognizes that agreement cannot always be reached.

In 1999, Texas enacted the so-called Futility Care Act. Other states, such as Virginia, Maryland, and California, have also enacted such laws that provide physicians a "safe harbor" from liability if they refuse a patient's or family's request for life-sustaining interventions. For instance, in Texas, when a disagreement about terminating interventions between the medical team and the family has not been resolved by an ethics consultation, the physician is tasked with trying to facilitate transfer of the patient to an institution willing to provide treatment. If this fails after 10 days, the hospital and physician may unilaterally withdraw treatments determined to be futile. The family may appeal to a state court. Early data suggest that the law increases futility consultations for the ethics committee and that, although most families concur with withdrawal, ~10–15% of families refuse to withdraw treatment. As of 2007, there had been 974 ethics committee consultations on medical futility cases and 65 in which committees ruled against families and gave notice that treatment would be terminated. In 2007,

a survey of Texas hospitals showed that 30% of hospitals had used the futility law in 213 adult cases and 42 pediatric cases. Treatment was withdrawn for 27 of those patients, and the remainder were transferred to other facilities or died while awaiting transfer.

EUTHANASIA AND PHYSICIAN-ASSISTED SUICIDE

Euthanasia and PAS are defined in **Table 12-8**. Terminating life-sustaining care and providing opioid medications to manage symptoms such as pain or dyspnea have long been considered ethical by the medical profession and legal by courts and should not be conflated with euthanasia or PAS.

Legal Aspects Euthanasia and PAS are legal in the Netherlands, Belgium, Luxembourg, Colombia, Canada, Spain, Western Australia, and New Zealand. Euthanasia was legalized in the Northern Territory of Australia in 1996, but that legislation was repealed 9 months later in 1997. Under certain conditions, a layperson in Switzerland or Germany can legally elect assisted suicide. In the United States, PAS is legal in Washington, D.C., and 10 states: Oregon, Washington State, Montana, Vermont, California, Colorado, Hawaii, Maine, New Jersey, and New Mexico. No state in the United States has legalized euthanasia. In the United States, multiple criteria must be met for PAS: the patient must have a terminal condition of <6 months and must be determined eligible through a process that includes a 15-day waiting period. In 2009, the state supreme court of Montana ruled that state law permits PAS for terminally ill patients. Many other countries, such as Portugal, are actively debating the legalization of euthanasia and/or PAS.

Practices Fewer than 10–20% of terminally ill patients actually consider euthanasia and/or PAS for themselves. Use of euthanasia and PAS is increasing but remains relatively rare. In all countries, even the Netherlands and Belgium where these practices have been tolerated and legal for many years, <5% of death occur by euthanasia or PAS. As of the most recent data, 4.7% of all deaths were by euthanasia or PAS in the Netherlands (2015) and 4.6% in Belgium (2013). Just 0.50% of all deaths in Oregon in 2019 (188 of 37,397 deaths) and 0.36% of all deaths in Washington State in 2018 (203 of 56,913 deaths) were reported to be by PAS, although these may be underestimates since the cause of some deaths of patients who received medications could not be verified.

In Belgium, the Netherlands, Oregon, and Washington, >70% of patients utilizing these interventions are dying of cancer; <10% of deaths by euthanasia or PAS involve patients with AIDS or amyotrophic

lateral sclerosis. While the numbers are small, in the Netherlands, the numbers of euthanasia or PAS cases in patients with psychiatric disorders, dementia, and the accumulation of health issues are increasing.

Pain is not the primary motivator for patients' requests for or interest in euthanasia and/or PAS. Among the first patients to receive PAS in Oregon, only 1 of the 15 patients had inadequate pain control, compared with 15 of the 43 patients in a control group who experienced inadequate pain relief. About 33% of patients in Oregon seeking PAS currently cite pain or fear of pain as their main reason for doing so. Conversely, depression and hopelessness are strongly associated with patient interest in euthanasia and PAS. Concerns about loss of dignity or autonomy or being a burden on family members appear to be more important factors motivating a desire for euthanasia or PAS. Losing autonomy (87% Oregon [OR], 85% Washington [WA]), not being able to enjoy activities (90% OR, 84% WA), and fear of losing dignity (72% OR, 69% WA) are the most-cited end-of-life concerns in both states. A high percentage of patients seeking PAS note being a burden on family (59% OR, 51% WA). A study from the Netherlands showed that depressed terminally ill cancer patients were four times more likely to request euthanasia and confirmed that uncontrolled pain was not associated with greater interest in euthanasia.

Euthanasia and PAS are no guarantee of a painless, quick death. Data from the Netherlands indicate that in as many as 20% of euthanasia and PAS cases technical and other problems arose, including patients waking from coma, not becoming comatose, regurgitating medications, and experiencing a prolonged time to death. Data from Oregon between 1998 and 2017 and Washington between 2009 and 2017 indicate that of patients who received PAS prescriptions, 81% died at home and prescribers were present in 9.7% of cases. The time between drug intake and coma ranged from 1 min to 11 h, and the time from drug intake to death ranged from 1 min to 104 h. The median time from ingestion to coma was 5 min and from ingestion to death was 25 min. In Oregon between 1998 and 2015, 53% of patients had no complications, 44% of patients had no data on complications, and 2.4% of patients had regurgitation after taking the prescribed medicine as the only complication. In addition, six patients awakened. In Washington State between 2014 and 2015, 1.4% of patients had regurgitation, one patient had a seizure, and the reported range of time to death extended to 30 h. In the Netherlands, problems were significantly more common in PAS, sometimes requiring the physician to intervene and provide euthanasia.

Regardless of whether they practice in a setting where euthanasia is legal or not, many physicians over the course of their careers will receive a patient request for euthanasia or PAS. In the United States, 18% of physicians have received a request for PAS and 11% have received a request for euthanasia. Three percent complied with a request for PAS, while 5% complied with a request for euthanasia. In the Netherlands, where the practices are legal, 77% of physicians have received a request for PAS or euthanasia and 60% have performed these interventions.

Competency in dealing with such a request is crucial. Although challenging, the request can also provide a chance to address intense suffering. After receiving a request for euthanasia and/or PAS, health care providers should carefully clarify the request with empathetic, open-ended questions to help elucidate the underlying cause for the request, such as, "What makes you want to consider this option?" Endorsing either moral opposition or moral support for the act tends to be counterproductive, giving an impression of being judgmental or of endorsing the idea that the patient's life is worthless. Health care providers must reassure the patient of continued care and commitment. The patient should be educated about alternative, less laden options, such as symptom management and withdrawing any unwanted treatments, and the reality of euthanasia and/or PAS, since the patient may have misconceptions about their effectiveness as well as the legal implications of the choice. Depression, hopelessness, and other symptoms of psychological distress, as well as physical suffering and economic burdens, are likely factors motivating the request, and such factors should be assessed and treated aggressively. After these interventions and clarification of options, most patients proceed with another approach,

TABLE 12-8 Definitions of Physician-Assisted Suicide and Euthanasia

TERM	DEFINITION	LEGAL STATUS
Voluntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death with the patient's informed consent	Netherlands, Belgium, Luxembourg, Canada, Colombia, Spain, Western Australia, New Zealand
Involuntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death when the patient was competent to consent but did not—e.g., the patient may not have been asked	Nowhere
Passive euthanasia	Withholding or withdrawing life-sustaining medical treatments from a patient to let him or her die (terminating life-sustaining treatments)	Everywhere
Physician-assisted suicide	A physician provides medications or other interventions to a patient with the understanding that the patient can use them to commit suicide	Netherlands, Belgium, Luxembourg, Canada, Colombia, Germany, Switzerland, Oregon, Washington, Montana, Vermont, California, Colorado, District of Columbia, Hawaii, Maine, New Jersey, New Mexico

declining life-sustaining interventions, possibly including refusal of nutrition and hydration.

CARE DURING THE LAST HOURS

Most laypersons have limited experiences with the actual dying process and death. They frequently do not know what to expect of the final hours and afterward. The family and other caregivers must be prepared, especially if the plan is for the patient to die at home.

Patients in the last days of life typically experience extreme weakness and fatigue and become bedbound; this can lead to pressure sores. The issue of turning patients who are near the end of life, however, must be balanced against the potential discomfort that movement may cause. Patients stop eating and drinking with drying of mucosal membranes and dysphagia. Careful attention to oral swabbing, lubricants for lips, and use of artificial tears can provide a form of care to substitute for attempts at feeding the patient. With loss of the gag reflex and dysphagia, patients may also experience accumulation of oral secretions, producing noises during respiration sometimes called “the death rattle.” Scopolamine can reduce the secretions. Patients also experience changes in respiration with periods of apnea or Cheyne-Stokes

breathing. Decreased intravascular volume and cardiac output cause tachycardia, hypotension, peripheral coolness, and livedo reticularis (skin mottling). Patients can have urinary and, less frequently, fecal incontinence. Changes in consciousness and neurologic function generally lead to two different paths to death.

Each of these terminal changes can cause patients and families distress, requiring reassurance and targeted interventions (**Table 12-9**). Informing families that these changes might occur and providing them with an information sheet can help preempt problems and minimize distress. Understanding that patients stop eating because they are dying, not dying because they have stopped eating, can reduce family and caregiver anxiety. Similarly, informing the family and caregivers that the “death rattle” may occur and that it is not indicative of suffocation, choking, or pain can reduce their worry from the breathing sounds.

Families and caregivers may also feel guilty about stopping treatments, fearing that they are “killing” the patient. This may lead to demands for interventions, such as feeding tubes, that may be ineffective. In such cases, the physician should remind the family and caregivers about the inevitability of events and the palliative goals.

TABLE 12-9 Managing Changes in the Patient’s Condition during the Final Days and Hours

CHANGES IN THE PATIENT’S CONDITION	POTENTIAL COMPLICATION	FAMILY’S POSSIBLE REACTION AND CONCERN	ADVICE AND INTERVENTION
Profound fatigue	Bedbound with development of pressure ulcers that are prone to infection, malodor, and pain, and joint pain	Patient is lazy and giving up.	Reassure family and caregivers that terminal fatigue will not respond to interventions and should not be resisted. Use an air mattress if necessary.
Anorexia	None	Patient is giving up; patient will suffer from hunger and will starve to death.	Reassure family and caregivers that the patient is not eating because he or she is dying; not eating at the end of life does not cause suffering or death. Forced feeding, whether oral, parenteral, or enteral, does not reduce symptoms or prolong life.
Dehydration	Dry mucosal membranes (see below)	Patient will suffer from thirst and die of dehydration.	Reassure family and caregivers that dehydration at the end of life does not cause suffering because patients lose consciousness before any symptom distress. Intravenous hydration can worsen symptoms of dyspnea by pulmonary edema and peripheral edema as well as prolong the dying process.
Dysphagia	Inability to swallow oral medications needed for palliative care		Do not force oral intake. Discontinue unnecessary medications that may have been continued, including antibiotics, diuretics, antidepressants, and laxatives. If swallowing pills is difficult, convert essential medications (analgesics, antiemetics, anxiolytics, and psychotropics) to oral solutions, buccal, sublingual, or rectal administration.
“Death rattle”—noisy breathing		Patient is choking and suffocating.	Reassure the family and caregivers that this is caused by secretions in the oropharynx and the patient is not choking. Reduce secretions with scopolamine (0.2–0.4 mg SC q4h or 1–3 patches q3d). Reposition patient to permit drainage of secretions. Do not suction. Suction can cause patient and family discomfort and is usually ineffective.
Apnea, Cheyne-Stokes respirations, dyspnea		Patient is suffocating.	Reassure family and caregivers that unconscious patients do not experience suffocation or air hunger. Apneic episodes are frequently a premorbid change. Opioids or anxiolytics may be used for dyspnea. Oxygen is unlikely to relieve dyspneic symptoms and may prolong the dying process.
Urinary or fecal incontinence	Skin breakdown if days until death Potential transmission of infectious agents to caregivers	Patient is dirty, malodorous, and physically repellent.	Remind family and caregivers to use universal precautions. Frequent changes of bedclothes and bedding. Use diapers, urinary catheter, or rectal tube if diarrhea or high urine output.
Agitation or delirium	Day/night reversal Hurt self or caregivers	Patient is in horrible pain and going to have a horrible death.	Reassure family and caregivers that agitation and delirium do not necessarily connote physical pain. Depending on the prognosis and goals of treatment, consider evaluating for causes of delirium and modifying medications. Manage symptoms with haloperidol, chlorpromazine, diazepam, or midazolam.
Dry mucosal membranes	Cracked lips, mouth sores, and candidiasis can also cause pain. Odor	Patient may be malodorous, physically repellent.	Use baking soda mouthwash or saliva preparation q15–30 min. Use topical nystatin for candidiasis. Coat lips and nasal mucosa with petroleum jelly q60–90 min. Use ophthalmic lubricants q4h or artificial tears q30 min.

Interventions may prolong the dying process and cause discomfort. Physicians also should emphasize that withholding treatments is both legal and ethical and that the family members are not the cause of the patient's death. This reassurance may have to be provided multiple times.

Hearing and touch are said to be the last senses to stop functioning. Whether this is the case or not, families and caregivers can be encouraged to communicate with the dying patient. Encouraging them to talk directly to the patient, even if he or she is unconscious, and hold the patient's hand or demonstrate affection in other ways can be an effective way to channel their urge "to do something" for the patient.

When the plan is for the patient to die at home, the physician must inform the family and caregivers how to determine that the patient has died. The cardinal signs are cessation of cardiac function and respiration; the pupils become fixed; the body becomes cool; muscles relax; and incontinence may occur. Remind the family and caregivers that the eyes may remain open even after the patient has died.

The physician should establish a plan for who the family or caregivers will contact when the patient is dying or has died. Without a plan, family members may panic and call 911, unleashing a cascade of unwanted events, from arrival of emergency personnel and resuscitation to hospital admission. The family and caregivers should be instructed to contact the hospice (if one is involved), the covering physician, or the on-call member of the palliative care team. They should also be told that the medical examiner need not be called unless the state requires it for all deaths. Unless foul play is suspected, the health care team need not contact the medical examiner either.

Just after the patient dies, even the best-prepared family may experience shock and loss and be emotionally distraught. They need time to assimilate the event and be comforted. Health care providers are likely to find it meaningful to write a bereavement card or letter to the family. The purpose is to communicate about the patient, perhaps emphasizing the patient's virtues and the honor it was to care for the patient, and to express concern for the family's hardship. Some physicians attend the funerals of their patients. Although this is beyond any medical obligation, the presence of the physician can be a source of support to the grieving family and provides an opportunity for closure for the physician.

Death of a spouse is a strong predictor of poor health, and even mortality, for the surviving spouse. It may be important to alert the spouse's physician about the death so that he or she is aware of symptoms that might require professional attention.

FURTHER READING

EMANUEL E et al: Attitudes and practices of euthanasia and physician-assisted suicide in the United States, Canada, and Europe. *JAMA* 316:79, 2016.

KELLEY AS, MEIER DE: Palliative care—A shifting paradigm. *N Engl J Med* 363:781, 2010.

KELLEY AS et al: Hospice enrollment saves money for Medicare and improves care quality across a number of different lengths-of-stay. *Health Aff* 32:552, 2012.

KELLEY AS et al: Palliative care for the seriously ill. *N Engl J Med* 373:747, 2015.

MACK JW et al: Associations between end-of-life discussion characteristics and care received near death: A prospective cohort study. *J Clin Oncol* 30:4387, 2012.

MURRAY SA et al: Illness trajectories and palliative care. *BMJ* 330:1007, 2005.

NEUMAN P et al: Medicare per capita spending by age and service: New data highlights oldest beneficiaries. *Health Aff (Millwood)* 34:335, 2015.

NICHOLAS LH et al: Regional variation in the association between advance directives and end-of-life Medicare expenditures. *JAMA* 306:1447, 2011.

ORNSTEIN KA et al: Evaluation of racial disparities in hospice use and end-of-life treatment intensity in the REGARDS cohort. *JAMA Netw Open* 3(8):e2014639, 2020.

QUINN KL et al: Association of receipt of palliative care interventions with health care use, quality of life, and symptom burden among adults with chronic noncancer illness: A systematic review and meta-analysis. *JAMA* 324:1439, 2020.

TENO JM et al: Change in end-of-life care for medicare beneficiaries: Site of death, place of care, and health transitions in 2000, 2005, and 2009. *JAMA* 309:470, 2013.

TENO JM et al: Site of death, place of care, and health care transitions among US Medicare beneficiaries, 2000–2015. *JAMA* 320:264, 2018.

VAN DEN BEUKEN-VANEVERDINGEN MH et al: Update on prevalence of pain in patients with cancer: Systematic review and meta-analysis. *J Pain Symptom Manage* 51:1070, 2016.

WEBSITES

AMERICAN ACADEMY OF HOSPICE AND PALLIATIVE MEDICINE: www.aahpm.org

CENTER TO ADVANCE PALLIATIVE CARE: <http://www.capc.org>

EDUCATION IN PALLIATIVE AND END OF LIFE CARE (EPEC): <http://www.epec.net>

FAMILY CAREGIVER ALLIANCE: <http://www.caregiver.org>

NATIONAL HOSPICE AND PALLIATIVE CARE ORGANIZATION (including state-specific advance directives): <http://www.nhpco.org>

NCCN: The National Comprehensive Cancer Network palliative care guidelines: <http://www.nccn.org>

OUR CARE WISHES ADVANCE CARE PLANNING TOOL: <https://www.ourcarewishes.org>

This page intentionally left blank

Section 1 Pain

13

Pain: Pathophysiology and Management

James P. Rathmell, Howard L. Fields



The province of medicine is to preserve and restore health and to relieve suffering. Understanding pain is essential to both of these goals. Because pain is universally understood as a signal of disease, it is the most common symptom that brings a patient to a physician's attention. The function of the pain sensory system is to protect the body and maintain homeostasis. It does this by detecting, localizing, and identifying potential or actual tissue-damaging processes. Because different diseases produce characteristic patterns of tissue damage, the quality, time course, and location of a patient's pain lend important diagnostic clues. It is the physician's responsibility to assess each patient promptly for any remediable cause underlying the pain and to provide rapid and effective pain relief whenever possible.

THE PAIN SENSORY SYSTEM

Pain is an unpleasant sensation localized to a part of the body. It is often described in terms of a penetrating or tissue-destructive process (e.g., stabbing, burning, twisting, tearing, squeezing) and/or of a bodily or emotional reaction (e.g., terrifying, nauseating, sickening). Furthermore, any pain of moderate or higher intensity is accompanied by anxiety and the urge to escape or terminate the feeling. These properties illustrate the duality of pain: it is both sensation and emotion. When it is acute, pain is characteristically associated with behavioral arousal and a stress response consisting of increased blood pressure, heart rate, pupil diameter, and plasma cortisol levels. In addition, local muscle contraction (e.g., limb flexion, abdominal wall rigidity) is often present.

PERIPHERAL MECHANISMS

The Primary Afferent Nociceptor A peripheral nerve consists of the axons of three different types of neurons: primary sensory afferents, motor neurons, and sympathetic postganglionic neurons (Fig. 13-1). The cell bodies of primary sensory afferents are located in the dorsal root ganglia within the vertebral foramina. The primary afferent axon has two branches: one projects centrally into the spinal cord and the other projects peripherally to innervate tissues. Primary afferents are classified by their diameter, degree of myelination, and conduction velocity. The largest diameter afferent fibers, A-beta ($A\beta$), respond maximally to light touch and/or moving stimuli; they are present primarily in nerves that innervate the skin. In normal individuals, the activity of these fibers does not produce pain. There are two other classes of primary afferent nerve fibers: the small diameter myelinated A-delta ($A\delta$) and the unmyelinated (C) axons (Fig. 13-1). These fibers are present in nerves to the skin and to deep somatic and visceral structures. Some tissues, such as the cornea, are innervated only by $A\delta$ and C fiber afferents.

Most $A\delta$ and C fiber afferents respond maximally to intense (painful) stimuli and produce the subjective experience of pain when they are activated; this defines them as *primary afferent nociceptors (pain receptors)*. The ability to detect painful stimuli is completely abolished when conduction in $A\delta$ and C fiber axons is blocked.

Individual primary afferent nociceptors can respond to several different types of noxious stimuli. For example, most nociceptors respond to heat; intense cold; intense mechanical distortion, such as a pinch; changes in pH, particularly an acidic environment; and application of chemical irritants including adenosine triphosphate (ATP), serotonin, bradykinin (BK), and histamine. The transient receptor potential cation channel subfamily V member 1 (TrpV1), also known as the vanilloid receptor, mediates perception of some noxious stimuli, especially heat sensations, by nociceptive neurons; it is activated by heat, acidic pH, endogenous mediators, and capsaicin, a component of hot chili peppers.

Sensitization When intense, repeated, or prolonged stimuli are applied to damaged or inflamed tissues, the threshold for activating primary afferent nociceptors is lowered, and the frequency of firing is higher for all stimulus intensities. Inflammatory mediators such as BK, nerve-growth factor, some prostaglandins (PGs), and leukotrienes contribute to this process, which is called *sensitization*. Sensitization occurs at the level of the peripheral nerve terminal (*peripheral sensitization*) as well as at the level of the dorsal horn of the spinal cord (*central sensitization*). Peripheral sensitization occurs in damaged or inflamed tissues, when inflammatory mediators activate intracellular signal transduction in nociceptors, prompting an increase in the production, transport, and membrane insertion of chemically gated and voltage-gated ion channels. These changes increase the excitability of nociceptor terminals and lower their threshold for activation by mechanical, thermal, and chemical stimuli. Central sensitization occurs when activity, generated by nociceptors during inflammation, enhances the excitability of nerve cells in the dorsal horn of the spinal cord. Following injury and resultant sensitization, normally innocuous stimuli can produce pain (termed *allodynia*). Sensitization is a clinically important process that contributes to tenderness, soreness, and *hyperalgesia* (increased pain intensity in response to the same noxious stimulus; e.g., pinprick causes severe pain). A striking example of sensitization is sunburned skin, in which severe pain can be produced by a gentle slap or a warm shower.

Sensitization is of particular importance for pain and tenderness in deep tissues. Viscera are normally relatively insensitive to noxious mechanical and thermal stimuli, although hollow viscera do generate

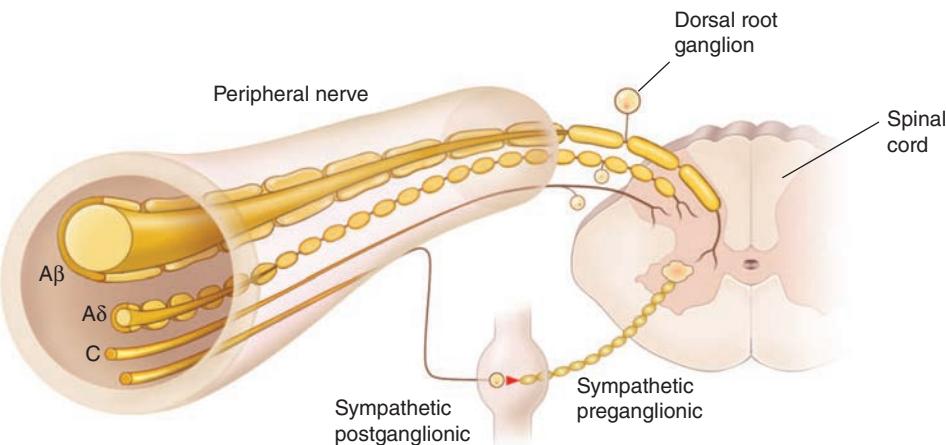


FIGURE 13-1 Components of a typical cutaneous nerve. There are two distinct functional categories of axons: primary afferents with cell bodies in the dorsal root ganglion and sympathetic postganglionic fibers with cell bodies in the sympathetic ganglion. Primary afferents include those with large-diameter myelinated ($A\beta$), small-diameter myelinated ($A\delta$), and unmyelinated (C) axons. All sympathetic postganglionic fibers are unmyelinated.

significant discomfort when distended. In contrast, when affected by a disease process with an inflammatory component, deep structures such as joints or hollow viscera characteristically become exquisitely sensitive to mechanical stimulation.

A large proportion of A_δ and C fiber afferents innervating viscera are completely insensitive in normal noninjured, noninflamed tissue. That is, they cannot be activated by known mechanical or thermal stimuli and are not spontaneously active. However, in the presence of inflammatory mediators, these afferents become sensitive to mechanical stimuli. Such afferents have been termed *silent nociceptors*, and their characteristic properties may explain how, under pathologic conditions, the relatively insensitive deep structures can become the source of severe and debilitating pain and tenderness. Low pH, PGs, leukotrienes, and other inflammatory mediators such as BK play a significant role in sensitization.

Nociceptor-Induced Inflammation Primary afferent nociceptors are not simply passive messengers of threats to tissue injury but also play an active role in tissue protection through a neuroeffector function. Most nociceptors contain polypeptide mediators, including substance P, calcitonin gene related peptide (CGRP), and cholecystokinin, that are released from their peripheral terminals when they are activated (Fig. 13-2). Substance P is an 11-amino-acid peptide that is released in peripheral tissues from primary afferent nociceptors and has multiple biologic activities. It is a potent vasodilator, causes mast cell degranulation, is a chemoattractant for leukocytes, and increases the production and release of inflammatory mediators. Interestingly, depletion of substance P from joints reduces the severity of experimental arthritis.

CENTRAL MECHANISMS

The Spinal Cord and Referred Pain The axons of primary afferent nociceptors enter the spinal cord via the dorsal root. They terminate in the dorsal horn of the spinal gray matter (Fig. 13-3). The terminals of primary afferent axons contact spinal neurons that transmit the pain signal to brain sites involved in pain perception. When primary afferents are activated by noxious stimuli, they release neurotransmitters from their terminals that excite the spinal cord neurons. The major neurotransmitter released is glutamate, which rapidly excites the second-order dorsal horn neurons. Primary afferent nociceptor terminals also release substance P and CGRP, which produce a slower and longer-lasting excitation of the dorsal horn neurons. The axon of each primary afferent contacts many spinal neurons, and each spinal neuron receives convergent inputs from many primary afferents.

The convergence of sensory inputs to a single spinal pain-transmission neuron is of great importance because it underlies the phenomenon of referred pain. All spinal neurons that receive input from the viscera and deep musculoskeletal structures also receive input from the skin. The convergence patterns are determined by the spinal segment of the dorsal root ganglion that supplies the afferent innervation of a structure. For example, the afferents that supply the central diaphragm are derived from the third and fourth cervical dorsal root ganglia. Primary afferents with cell bodies in these same ganglia supply the skin of the shoulder and lower neck. Thus, sensory inputs from both the shoulder skin and the central diaphragm converge on pain-transmission neurons in the third and fourth cervical spinal segments. *Because of this convergence and the fact that the spinal neurons are most often activated by inputs from the skin, activity evoked in spinal neurons by input from deep structures is often mislocalized by the patient to a bodily location that roughly corresponds with the region of skin innervated by the same spinal segment.* Thus, inflammation near the central diaphragm is often reported as shoulder discomfort. This spatial displacement of pain sensation from the site of the injury that produces it is known as *referred pain*.

Ascending Pathways for Pain A majority of spinal neurons contacted by primary afferent nociceptors send their axons to the contralateral thalamus. These axons form the contralateral spinothalamic tract, which lies in the anterolateral white matter of the spinal cord,

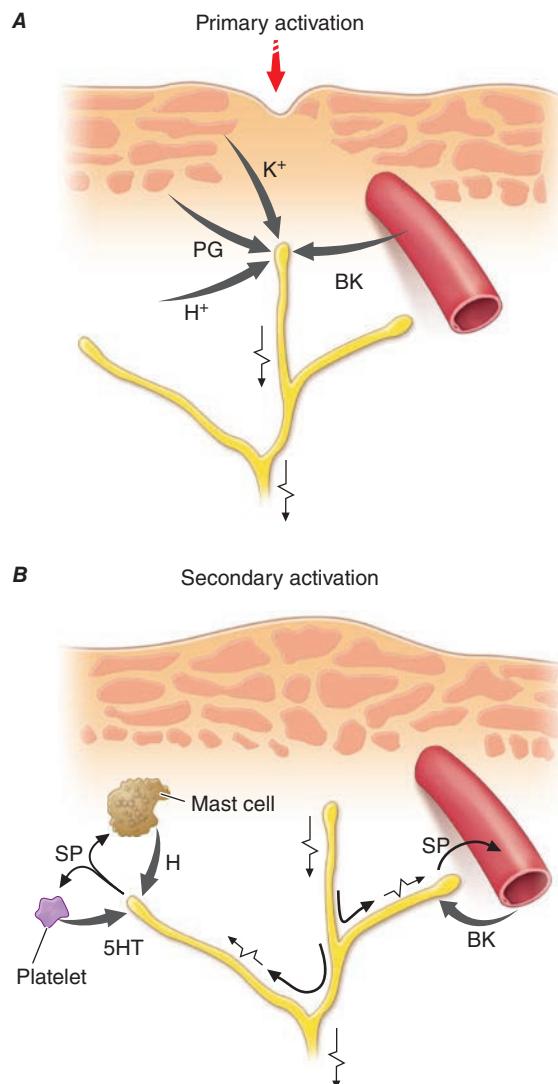


FIGURE 13-2 Events leading to activation, sensitization, and spread of sensitization of primary afferent nociceptor terminals. **A.** Direct activation by intense pressure and consequent cell damage. Cell damage induces lower pH (H^+) and leads to release of potassium (K^+) and to synthesis of prostaglandins (PGs) and bradykinin (BK). PGs increase the sensitivity of the terminal to BK and other pain-producing substances. **B.** Secondary activation. Impulses generated in the stimulated terminal propagate not only to the spinal cord but also into other terminal branches where they induce the release of peptides, including substance P (SP). Substance P causes vasodilation and neurogenic edema with further accumulation of BK. Substance P also causes the release of histamine (H) from mast cells and serotonin (5HT) from platelets.

the lateral edge of the medulla, and the lateral pons and midbrain. The spinothalamic pathway is crucial for pain sensation in humans. Interruption of this pathway produces permanent deficits in pain and temperature discrimination.

Spinothalamic tract axons ascend to several regions of the thalamus. There is tremendous divergence of the pain signal from these thalamic sites to several distinct areas of the cerebral cortex that subserve different aspects of the pain experience (Fig. 13-4). One of the thalamic projections is to the somatosensory cortex. This projection mediates the sensory discriminative aspects of pain, i.e., its location, intensity, and quality. Other thalamic neurons project to cortical regions that are linked to emotional responses, such as the cingulate and insular cortex. These pathways to the frontal cortex subserve the affective or unpleasant emotional dimension of pain. This affective dimension of pain produces suffering and exerts potent control of behavior. Because of this dimension, fear is a constant companion of pain. As a consequence, injury or surgical lesions to areas of the frontal cortex activated by painful stimuli can diminish the emotional impact of pain while

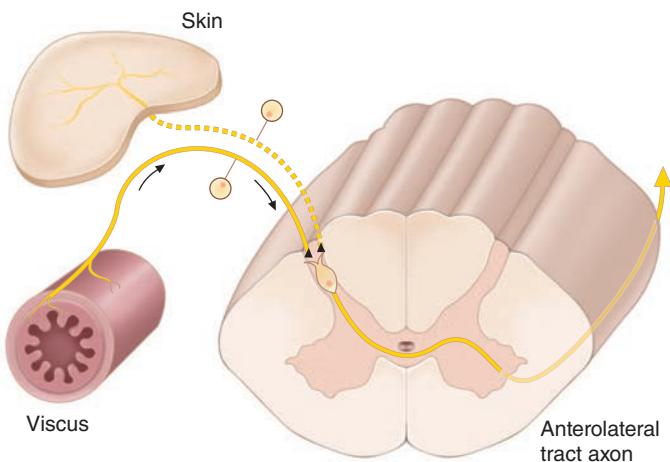


FIGURE 13-3 The convergence-projection hypothesis of referred pain. According to this hypothesis, visceral afferent nociceptors converge on the same pain-projection neurons as the afferents from the somatic structures in which the pain is perceived. The brain has no way of knowing the actual source of input and mistakenly “projects” the sensation to the somatic structure.

largely preserving the individual’s ability to recognize noxious stimuli as painful.

PAIN MODULATION

The pain produced by injuries of similar magnitude is remarkably variable in different situations and in different individuals. For example, athletes have been known to sustain serious fractures with only minor pain, and Beecher’s classic World War II survey revealed that many soldiers in battle were unbothered by injuries that would have produced agonizing pain in civilian patients. Furthermore, even the suggestion

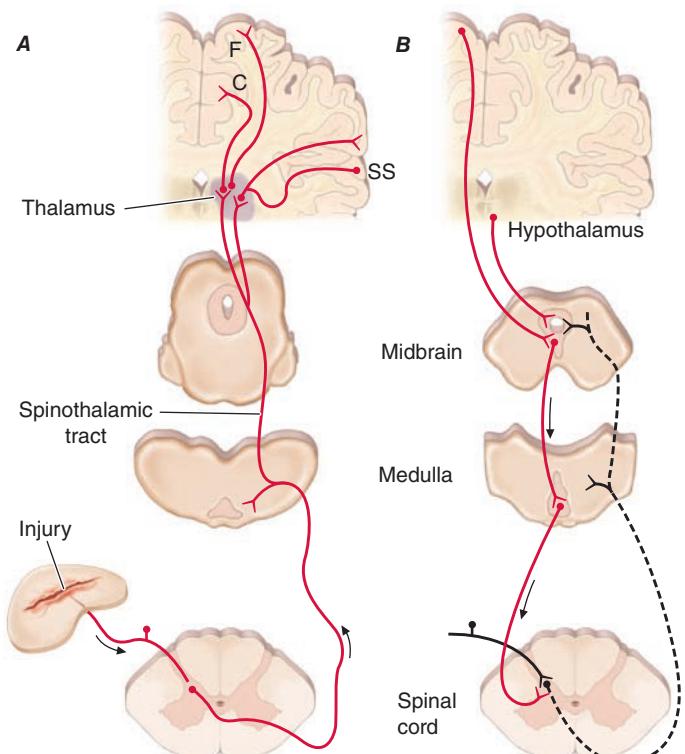


FIGURE 13-4 Pain-transmission and modulatory pathways. **A.** Transmission system for nociceptive messages. Noxious stimuli activate the sensitive peripheral ending of the primary afferent nociceptor by the process of transduction. The message is then transmitted over the peripheral nerve to the spinal cord, where it synapses with cells of origin of the major ascending pain pathway, the spinothalamic tract. The message is relayed in the thalamus to the anterior cingulate (C), frontal insular (F), and somatosensory cortex (SS). **B.** Pain-modulation network. Inputs from frontal cortex and hypothalamus activate cells in the midbrain that control spinal pain-transmission cells via cells in the medulla.

that a treatment will relieve pain can have a significant analgesic effect (the *placebo effect*). On the other hand, many patients find even minor injuries such as venipuncture frightening and unbearable, and the expectation of pain can induce pain even without a noxious stimulus. The suggestion that pain will worsen following administration of an inert substance can increase its perceived intensity (the *nocebo effect*).

The powerful effect of expectation and other psychological variables on the perceived intensity of pain is explained by brain circuits that modulate the activity of the pain-transmission pathways. One of these circuits has links to the hypothalamus, midbrain, and medulla, and it selectively controls spinal pain-transmission neurons through a descending pathway (Fig. 13-4).

Human brain-imaging studies have implicated this pain-modulating circuit in the pain-relieving effect of attention, suggestion, and opioid analgesic medications (Fig. 13-5). Furthermore, each of the component structures of the pathway contains opioid receptors and is sensitive to the direct application of opioid drugs. In animals, lesions of this descending modulatory system reduce the analgesic effect of

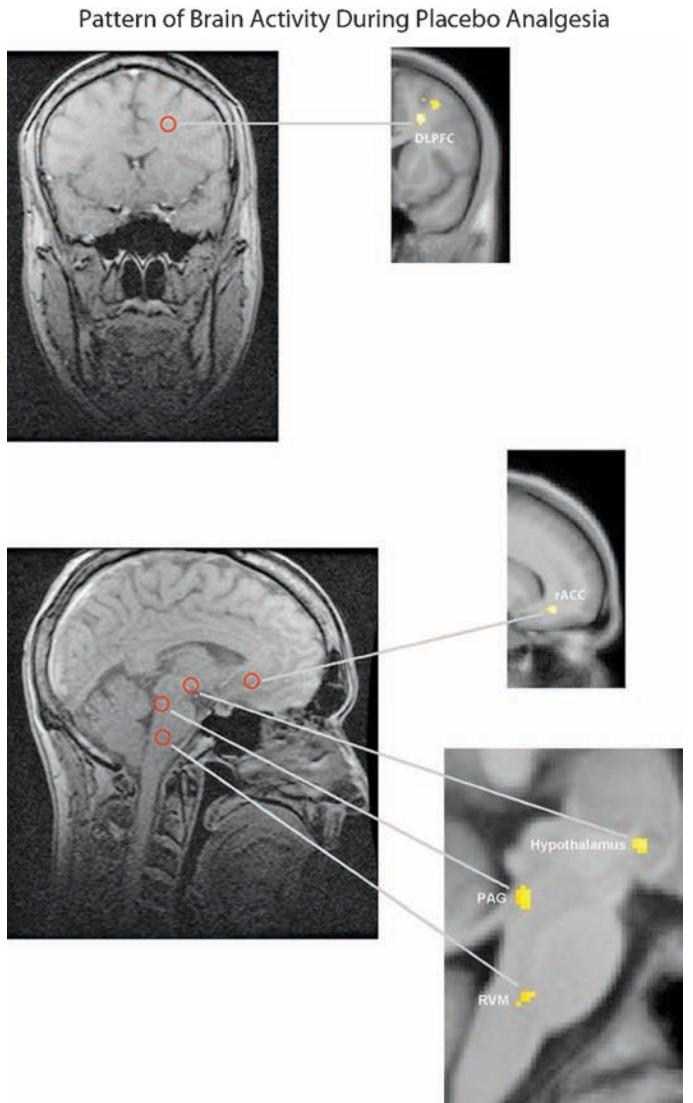


FIGURE 13-5 Functional magnetic resonance imaging (fMRI) demonstrates placebo-enhanced brain activity in anatomic regions correlating with the opioidergic descending pain control system. **Top panel:** Frontal fMRI image shows placebo-enhanced brain activity in the dorsal lateral prefrontal cortex (DLPFC). **Bottom panel:** Sagittal fMRI images show placebo-enhanced responses in the rostral anterior cingulate cortex (rACC), the rostral ventral medullae (RVM), the periaqueductal gray (PAG) area, and the hypothalamus. The placebo-enhanced activity in all areas was reduced by naloxone, demonstrating the link between the descending opioidergic system and the placebo analgesic response. (F Eippert et al: Activation of the opioidergic descending pain control system underlies placebo analgesia. *Neuron* 63(4):533-543, 2009.)

systemically administered opioids such as morphine. Along with the opioid receptor, the component nuclei of this pain-modulating circuit contain endogenous opioid peptides such as the enkephalins and β -endorphin.

The most reliable way to activate this endogenous opioid-mediated modulating system is by suggestion of pain relief or by intense emotion directed away from the pain-causing injury (e.g., during severe threat or an athletic competition). In fact, pain-relieving endogenous opioids are released following surgical procedures and in patients given a placebo for pain relief.

Pain-modulating circuits can enhance as well as suppress pain. Both pain-inhibiting and pain-facilitating neurons in the medulla project to and control spinal pain-transmission neurons. Because pain-transmission neurons can be activated by modulatory neurons, it is theoretically possible to generate a pain signal with no peripheral noxious stimulus. In fact, human functional imaging studies have demonstrated increased activity in this circuit during migraine headaches. A central circuit that facilitates pain could account for the finding that pain can be induced by suggestion or enhanced by expectation and provides a framework for understanding how psychological factors can contribute to chronic pain.

■ NEUROPATHIC PAIN

Lesions of the peripheral or central nociceptive pathways typically result in a loss or impairment of pain sensation. Paradoxically, damage to or dysfunction of these pathways can also produce pain. For example, damage to peripheral nerves, as occurs in diabetic neuropathy, or to primary afferents, as in herpes zoster infection, can result in pain that is referred to the body region innervated by the damaged nerves. Pain may also be produced by damage to the central nervous system (CNS), for example, in some patients following trauma or vascular injury to the spinal cord, brainstem, or thalamic areas that contain central nociceptive pathways. Such pains are termed *neuropathic* and are often severe and resistant to standard treatments for pain.

Neuropathic pain typically has an unusual burning, tingling, or electric shock-like quality and may occur spontaneously, without any stimulus, or be triggered by very light touch. These features are rare in other types of pain. On examination, a sensory deficit is characteristically co-extensive with the area of the patient's pain. *Hyperpathia*, a greatly exaggerated pain response to innocuous or mild nociceptive stimuli, especially when applied repeatedly, is also characteristic of neuropathic pain; patients often complain that the very lightest moving stimulus evokes exquisite pain (allodynia). In this regard, it is of clinical interest that a topical preparation of 5% lidocaine in patch form is effective for patients with postherpetic neuralgia who have prominent allodynia.

A variety of mechanisms contribute to neuropathic pain. As with sensitized primary afferent nociceptors, damaged primary afferents, including nociceptors, become highly sensitive to mechanical stimulation and may generate impulses in the absence of stimulation. Increased sensitivity and spontaneous activity are due, in part, to an increased density of sodium channels in the damaged nerve fiber. Damaged primary afferents may also develop sensitivity to norepinephrine. Interestingly, spinal cord pain-transmission neurons cut off from their normal input may also become spontaneously active. Thus, both central and peripheral nervous system hyperactivity contribute to neuropathic pain.

Sympathetically Maintained Pain Patients with peripheral nerve injury occasionally develop spontaneous pain in or beyond the region innervated by the nerve. This pain is often described as having a burning quality. The pain typically begins after a delay of hours to days or even weeks and is accompanied by swelling of the extremity, periarticular bone loss, and arthritic changes in the distal joints. Early in the course of the condition, the pain may be relieved by a local anesthetic block of the sympathetic innervation to the affected extremity. Damaged primary afferent nociceptors acquire adrenergic sensitivity and can be activated by stimulation of the sympathetic outflow. This constellation of spontaneous pain and signs of sympathetic dysfunction following injury has been termed *complex regional pain*

syndrome (CRPS). When this occurs after an identifiable nerve injury, it is termed CRPS type II (also known as posttraumatic neuralgia or, if severe, *causalgia*). When a similar clinical picture appears without obvious nerve injury, it is termed CRPS type I (also known as *reflex sympathetic dystrophy*). CRPS can be produced by a variety of injuries, including fractures of bone, soft tissue trauma, myocardial infarction, and stroke. CRPS type I typically resolves with symptomatic treatment; however, when it persists, detailed examination often reveals evidence of peripheral nerve injury. Although the pathophysiology of CRPS is poorly understood, the pain and the signs of inflammation, when acute, can be rapidly relieved by blocking the sympathetic nervous system. This implies that sympathetic activity can activate undamaged nociceptors when inflammation is present. Signs of sympathetic hyperactivity should be sought in patients with posttraumatic pain and inflammation and no other obvious explanation.

TREATMENT

Acute Pain

The ideal treatment for any pain is to remove the cause; thus, while treatment can be initiated immediately, efforts to establish the underlying etiology should always proceed as treatment begins. Sometimes, treating the underlying condition does not immediately relieve pain. Furthermore, some conditions are so painful that rapid and effective analgesia is essential (e.g., the postoperative state, burns, trauma, cancer, or sickle cell crisis). Analgesic medications are a first line of treatment in these cases, and all practitioners should be familiar with their use.

ASPIRIN, ACETAMINOPHEN, AND NONSTEROIDAL ANTI-INFLAMMATORY AGENTS (NSAIDS)

These drugs are considered together because they are used for similar problems and may have a similar mechanism of action (**Table 13-1**). All these compounds inhibit cyclooxygenase (COX), and except for acetaminophen, all have anti-inflammatory actions, especially at higher dosages. They are particularly effective for mild to moderate headache and for pain of musculoskeletal origin.

Because they are effective for these common types of pain and are available without prescription, COX inhibitors are by far the most commonly used analgesics. They are absorbed well from the gastrointestinal tract and, with occasional use, have only minimal side effects. With chronic use, gastric irritation is a common side effect of aspirin and NSAIDs and is the problem that most frequently limits the dose that can be given. Gastric irritation is most severe with aspirin, which may cause erosion and ulceration of the gastric mucosa leading to bleeding or perforation. Because aspirin irreversibly acetylates platelet COX and thereby interferes with coagulation of the blood, gastrointestinal bleeding is a particular risk. Older age and history of gastrointestinal disease increase the risks of aspirin and NSAIDs. In addition to the well-known gastrointestinal toxicity of NSAIDs, nephrotoxicity is a significant problem for patients using these drugs on a chronic basis. Patients at risk for renal insufficiency, particularly those with significant contraction of their intravascular volume as occurs with chronic diuretic use or acute hypovolemia, should avoid NSAIDs. NSAIDs can also increase blood pressure in some individuals. Long-term treatment with NSAIDs requires regular blood pressure monitoring and treatment if necessary. Although toxic to the liver when taken in high doses, acetaminophen rarely produces gastric irritation and does not interfere with platelet function.

The introduction of parenteral forms of NSAIDs, ketorolac and diclofenac, extends the usefulness of this class of compounds in the management of acute severe pain. Both agents are sufficiently potent and rapid in onset to supplant opioids as first-line treatment for many patients with acute severe headache and musculoskeletal pain.

There are two major classes of COX: COX-1 is constitutively expressed, and COX-2 is induced in the inflammatory state.

TABLE 13-1 Drugs for Relief of Pain

GENERIC NAME	DOSE, mg	INTERVAL	COMMENTS					
Nonnarcotic Analgesics: Usual Doses and Intervals								
Acetylsalicylic acid	650 PO	q4h	Enteric-coated preparations available					
Acetaminophen	650 PO	q4h	Side effects uncommon					
Ibuprofen	400 PO	q4–6h	Available without prescription					
Naproxen	250–500 PO	q12h	Naproxen is the common NSAID that poses the least cardiovascular risk, but it has a somewhat higher incidence of gastrointestinal bleeding					
Fenoprofen	200 PO	q4–6h	Contraindicated in renal disease					
Indomethacin	25–50 PO	q8h	Gastrointestinal side effects common					
Ketorolac	15–60 IM/IV	q4–6h	Available for parenteral use					
Celecoxib	100–200 PO	q12–24h	Useful for arthritis					
Valdecoxib	10–20 PO	q12–24h	Removed from U.S. market in 2005					
GENERIC NAME	PARENTERAL DOSE, mg	PO DOSE, mg	COMMENTS					
Narcotic Analgesics: Usual Doses and Intervals								
Codeine	30–60 q4h	30–60 q4h	Nausea common					
Oxycodone	—	5–10 q4–6h	Usually available with acetaminophen or aspirin					
Oxycodone extended-release	—	10–40 q12h	Oral extended-release tablet; high potential for misuse					
Morphine	5 q4h	30 q4h						
Morphine sustained release	—	15–60 bid to tid	Oral slow-release preparation					
Hydromorphone	1–2 q4h	2–4 q4h	Shorter acting than morphine sulfate					
Levorphanol	2 q6–8h	4 q6–8h	Longer acting than morphine sulfate; absorbed well PO					
Methadone	5–10 q6–8h	5–20 q6–8h	Due to long half-life, respiratory depression and sedation may persist after analgesic effect subsides; therapy should not be initiated with >40 mg/d, and dose escalation should be made no more frequently than every 3 days					
Meperidine	50–100 q3–4h	300 q4h	Poorly absorbed PO; normeperidine is a toxic metabolite; routine use of this agent is not recommended					
Butorphanol	—	1–2 q4h	Intranasal spray					
Fentanyl	25–100 µg/h	—	72-h transdermal patch					
Buprenorphine	5–20 µg/h	—	7-day transdermal patch					
Buprenorphine	0.3 q6–8h	—	Parenteral administration					
Tramadol	—	50–100 q4–6h	Mixed opioid/adrenergic action					
GENERIC NAME	UPTAKE BLOCKADE 5-HT NE	SEDATIVE POTENCY	ANTICHOLINERGIC POTENCY	ORTHOSTATIC HYPOTENSION	CARDIAC ARRHYTHMIA	AVERAGE DOSE, mg/d	RANGE, mg/d	
Antidepressants^a								
Doxepin	++	+	High	Moderate	Moderate	Less	200	75–400
Amitriptyline	++++	++	High	Highest	Moderate	Yes	150	25–300
Imipramine	++++	++	Moderate	Moderate	High	Yes	200	75–400
Nortriptyline	+++	++	Moderate	Moderate	Low	Yes	100	40–150
Desipramine	+++	++++	Low	Low	Low	Yes	150	50–300
Venlafaxine	+++	++	Low	None	None	No	150	75–400
Duloxetine	+++	+++	Low	None	None	No	40	30–60
GENERIC NAME	PO DOSE, mg	INTERVAL	COMMENTS					
Anticonvulsants and Antiarrhythmics^a								
Carbamazepine	200–300	q6h	Rare aplastic anemia, GI irritation, hepatotoxicity					
Oxcarbamazepine	300	bid	Similar to carbamazepine					
Gabapentin ^b	600–1200	q8h	Dizziness, GI irritation; useful in trigeminal neuralgia					
Pregabalin	150–600	bid	Similar to gabapentin; dry mouth, edema					

^aAntidepressants, anticonvulsants, and antiarrhythmics have not been approved by the U.S. Food and Drug Administration (FDA) for the treatment of pain. ^bGabapentin in doses up to 1800 mg/d is FDA approved for postherpetic neuralgia.

Abbreviations: 5-HT, serotonin; NE, norepinephrine; NSAID, nonsteroidal anti-inflammatory agent.

COX-2-selective drugs have similar analgesic potency and produce less gastric irritation than the nonselective COX inhibitors. The use of COX-2-selective drugs does not appear to lower the risk of nephrotoxicity compared to nonselective NSAIDs. On the other hand, COX-2-selective drugs offer a significant benefit in the management of acute postoperative pain because they do not affect blood coagulation. Nonselective COX inhibitors (especially

aspirin) are usually contraindicated postoperatively because they impair platelet-mediated blood clotting and are thus associated with increased bleeding at the operative site. COX-2 inhibitors, including celecoxib (Celebrex), are associated with increased cardiovascular risk, including cardiovascular death, myocardial infarction, stroke, heart failure, or a thromboembolic event. It appears that this is a class effect of NSAIDs, excluding aspirin. These drugs

are contraindicated in patients in the immediate period after coronary artery bypass surgery and should be used with caution in elderly patients and those with a history of or significant risk factors for cardiovascular disease.

OPIOID ANALGESICS

Opioids are the most potent pain-relieving drugs currently available. Of all analgesics, they have the broadest range of efficacy and provide the most reliable and effective treatment for rapid pain relief. Although side effects are common, most are reversible: nausea, vomiting, pruritus, sedation, and constipation are the most frequent and bothersome side effects. Respiratory depression is uncommon at standard analgesic doses but can be life-threatening. Opioid-related side effects can be reversed rapidly with the narcotic antagonist naloxone. Many physicians, nurses, and patients have a certain trepidation about using opioids that is based on a fear of initiating addiction in their patients. In fact, there is a very small chance of patients becoming addicted to narcotics as a result of their appropriate medical use. For chronic pain, particularly chronic noncancer pain, the risk of addiction in patients taking opioids on a chronic basis remains small, but the risk does appear to increase with dose escalation. The physician should not hesitate to use opioid analgesics in patients with acute severe pain. Table 13-1 lists the most commonly used opioid analgesics.

Opioids produce analgesia by actions in the CNS. They activate pain-inhibitory neurons and directly inhibit pain-transmission neurons. Most of the commercially available opioid analgesics act at the same opioid receptor (μ -receptor), differing mainly in potency, speed of onset, duration of action, and optimal route of administration. Some side effects are due to accumulation of nonopiod metabolites that are unique to individual drugs. One striking example of this is normeperidine, a metabolite of meperidine. At higher doses of meperidine, typically >1 g/d, accumulation of normeperidine can produce hyperexcitability and seizures that are not reversible with naloxone. Normeperidine accumulation is increased in patients with renal failure.

The most rapid pain relief is obtained by intravenous administration of opioids; relief with oral administration is significantly slower. Because of the potential for respiratory depression, patients with any form of respiratory compromise must be kept under close observation following opioid administration; an oxygen-saturation monitor may be useful, but only in a setting where the monitor is under constant surveillance. Opioid-induced respiratory depression is primarily manifest as a reduction in respiratory rate and is typically accompanied by sedation. A fall in oxygen saturation represents a critical level of respiratory depression and the need for immediate intervention to prevent life-threatening hypoxemia. Newer monitoring devices that incorporate capnography or pharyngeal air flow can detect apnea at the point of onset and should be used in hospitalized patients. Ventilatory assistance should be maintained until the opioid-induced respiratory depression has resolved. The opioid antagonist naloxone should be readily available whenever opioids are used at high doses or in patients with compromised pulmonary function. Opioid effects are dose-related, and there is great variability among patients in the doses that relieve pain and produce side effects. Synergistic respiratory depression is common when opioids are administered with other CNS depressants. Co-administration of benzodiazepines is particularly likely to produce respiratory depression and should be avoided, especially in outpatient pain management. Because of this variability in patient response, initiation of therapy requires titration to optimal dose and interval. The most important principle is to provide adequate pain relief. This requires determining whether the drug has adequately relieved the pain and timely reassessment to determine the optimal interval for dosing. *The most common error made by physicians in managing severe pain with opioids is to prescribe an inadequate dose. Because many patients are reluctant to complain, this practice leads to needless suffering.*

In the absence of sedation at the expected time of peak effect, a physician should not hesitate to repeat the initial dose to achieve satisfactory pain relief.

A now standard approach to the problem of achieving adequate pain relief is the use of patient-controlled analgesia (PCA). PCA uses a microprocessor-controlled infusion device that can deliver a baseline continuous dose of an opioid drug as well as preprogrammed additional doses whenever the patient pushes a button. The patient can then titrate the dose to the optimal level. This approach is used most extensively for the management of postoperative pain, but there is no reason why it should not be used for any hospitalized patient with persistent severe pain. PCA is also used for short-term home care of patients with intractable pain, such as that caused by metastatic cancer.

It is important to understand that the PCA device delivers small, repeated doses to maintain pain relief; in patients with severe pain, the pain must first be brought under control with a loading dose before transitioning to the PCA device. The bolus dose of the drug (typically 1 mg of morphine, 0.2 mg of hydromorphone, or 10 μ g of fentanyl) can then be delivered repeatedly as needed. To prevent overdosing, PCA devices are programmed with a lockout period after each demand dose is delivered (typically starting at 10 min) and a limit on the total dose delivered per hour. Although some have advocated the use of a simultaneous continuous or basal infusion of the PCA drug, this may increase the risk of respiratory depression and has not been shown to increase the overall efficacy of the technique.

The availability of new routes of administration has extended the usefulness of opioid analgesics. Most important is the availability of spinal administration. Opioids can be infused through a spinal catheter placed either intrathecally or epidurally. By applying opioids directly to the spinal or epidural space adjacent to the spinal cord, regional analgesia can be obtained using relatively low total doses. Indeed, the dose required to produce effective analgesia when using morphine intrathecally (0.1–0.3 mg) is a fraction of that required to produce similar analgesia when administered intravenously (5–10 mg). In this way, side effects such as sedation, nausea, and respiratory depression can be minimized. This approach has been used extensively during labor and delivery and for postoperative pain relief following surgical procedures. Continuous intrathecal delivery via implanted spinal drug-delivery systems is now commonly used, particularly for the treatment of cancer-related pain that would require sedating doses for adequate pain control if given systemically. Opioids can also be given intranasally (butorphanol), rectally, and transdermally (fentanyl and buprenorphine), or through the oral mucosa (fentanyl), thus avoiding the discomfort of frequent injections in patients who cannot be given oral medication. The fentanyl and buprenorphine transdermal patches have the advantage of providing fairly steady plasma levels, which may improve patient comfort.

Recent additions to the armamentarium for treating opioid-induced side effects are the peripherally acting opioid antagonists alvimopan (Entereg) and methylnaltrexone (Relistor). Alvimopan is available as an orally administered agent that is restricted to the intestinal lumen by limited absorption; methylnaltrexone is available in a subcutaneously administered form that has virtually no penetration into the CNS. Both agents act by binding to peripheral μ -receptors, thereby inhibiting or reversing the effects of opioids at these peripheral sites. The action of both agents is restricted to receptor sites outside of the CNS; thus, these drugs can reverse the adverse effects of opioid analgesics that are mediated through their peripheral receptors without reversing their CNS-mediated analgesic effects. Alvimopan has proven effective in lowering the duration of persistent ileus following abdominal surgery in patients receiving opioid analgesics for postoperative pain control. Methylnaltrexone has proven effective for relief of opioid-induced constipation in patients taking opioid analgesics on a chronic basis.

Opioid and COX Inhibitor Combinations When used in combination, opioids and COX inhibitors have additive effects. Because a lower dose of each can be used to achieve the same degree of pain relief and their side effects are nonadditive, such combinations are used to lower the severity of dose-related side effects. However, fixed-ratio combinations of an opioid with acetaminophen carry an important risk. Dose escalation as a result of increased severity of pain or decreased opioid effect as a result of tolerance may lead to ingestion of levels of acetaminophen that are toxic to the liver. Although acetaminophen-related hepatotoxicity is uncommon, it remains a significant cause for liver failure. Thus, many practitioners have moved away from the use of opioid-acetaminophen combination analgesics to avoid the risk of excessive acetaminophen exposure as the dose of the analgesic is escalated.

CHRONIC PAIN

Managing patients with chronic pain is intellectually and emotionally challenging. Sensitization of the nervous system can occur without an obvious precipitating cause, e.g., fibromyalgia, or chronic headache. In many patients, chronic pain becomes a distinct disease unto itself. The pain-generating mechanism is often difficult or impossible to determine with certainty; such patients are demanding of the physician's time and often appear emotionally distraught. The traditional medical approach of seeking an obscure organic pathology is often unhelpful. On the other hand, psychological evaluation and behaviorally based treatment paradigms are frequently helpful, particularly in the setting of a multidisciplinary pain-management center. Unfortunately, this approach, while effective, remains largely underused in current medical practice.

There are several factors that can cause, perpetuate, or exacerbate chronic pain. First, of course, the patient may simply have a disease that is characteristically painful for which there is presently no cure. Arthritis, cancer, chronic daily headaches, fibromyalgia, and diabetic neuropathy are examples of this. Second, there may be secondary perpetuating factors that are initiated by disease and persist after that disease has resolved. Examples include damaged sensory nerves, sympathetic efferent activity, and painful reflex muscle contraction (spasm). Finally, a variety of psychological conditions can exacerbate or even cause pain.

There are certain areas to which special attention should be paid in a patient's medical history. Because depression is the most common emotional disturbance in patients with chronic pain, patients should be questioned about their mood, appetite, sleep patterns, and daily activity. A simple standardized questionnaire, such as the Beck Depression Inventory, can be a useful screening device. It is important to remember that major depression is a common, treatable, and potentially fatal illness.

Other clues that a significant emotional disturbance is contributing to a patient's chronic pain complaint include pain that occurs in multiple, unrelated sites; a pattern of recurrent, but separate, pain problems beginning in childhood or adolescence; pain beginning at a time of emotional trauma, such as the loss of a parent or spouse; a history of physical or sexual abuse; and past or present substance abuse.

On examination, special attention should be paid to whether the patient guards the painful area and whether certain movements or postures are avoided because of pain. Discovering a mechanical component to the pain can be useful both diagnostically and therapeutically. Painful areas should be examined for deep tenderness, noting whether this is localized to muscle, ligamentous structures, or joints. Chronic myofascial pain is very common, and in these patients, deep palpation may reveal highly localized trigger points that are firm bands or knots in muscle. Relief of the pain following injection of local anesthetic into these trigger points supports the diagnosis. A neuropathic component to the pain is indicated by evidence of nerve damage, such as sensory impairment, exquisitely sensitive skin (allodynia), weakness, and muscle atrophy, or loss of deep tendon reflexes. Evidence suggesting sympathetic nervous system involvement includes the presence of diffuse

swelling, changes in skin color and temperature, and hypersensitive skin and joint tenderness compared with the normal side. Relief of the pain with a sympathetic block supports the diagnosis, but once the condition becomes chronic, the response to sympathetic blockade is of variable magnitude and duration; the role for repeated sympathetic blocks in the overall management of CRPS is unclear.

A guiding principle in evaluating patients with chronic pain is to assess both emotional and somatic causal and perpetuating factors before initiating therapy. Addressing these issues together, rather than waiting to address emotional issues after somatic causes of pain have been ruled out, improves compliance in part because it assures patients that a psychological evaluation does not mean that the physician is questioning the validity of their complaint. Even when a somatic cause for a patient's pain can be found, it is still wise to look for other factors. For example, a cancer patient with painful bony metastases may have additional pain due to nerve damage and may also be depressed. Optimal therapy requires that each of these factors be assessed and treated.

TREATMENT

Chronic Pain

Once the evaluation process has been completed and the likely causative and exacerbating factors identified, an explicit treatment plan should be developed. An important part of this process is to identify specific and realistic functional goals for therapy, such as getting a good night's sleep, being able to go shopping, or returning to work. A multidisciplinary approach that uses medications, counseling, physical therapy, nerve blocks, and even surgery may be required to improve the patient's quality of life. There are also some newer, minimally invasive procedures that can be helpful for some patients with intractable pain. These include image-guided interventions such as epidural injection of glucocorticoids for acute radicular pain and radiofrequency treatment of the facet joints for chronic facet-related back and neck pain. For patients with severe and persistent pain that is unresponsive to more conservative treatment, placement of electrodes on peripheral nerves or within the spinal canal on nerve roots or in the space overlying the dorsal columns of the spinal cord (spinal cord stimulation) or implantation of intrathecal drug-delivery systems has shown significant benefit. The criteria for predicting which patients will respond to these procedures continue to evolve. They are generally reserved for patients who have not responded to conventional pharmacologic approaches. Referral to a multidisciplinary pain clinic for a full evaluation should precede any invasive procedure. Such referrals are clearly not necessary for all chronic pain patients. For some, pharmacologic management alone can provide adequate relief.

ANTIDEPRESSANT MEDICATIONS

The tricyclic antidepressants (TCAs), particularly nortriptyline and desipramine (Table 13-1), are useful for the management of chronic pain. Although developed for the treatment of depression, the TCAs have a spectrum of dose-related biologic activities that include analgesia in a variety of chronic clinical conditions. Although the mechanism is unknown, the analgesic effect of TCAs has a more rapid onset and occurs at a lower dose than is typically required for the treatment of depression. Furthermore, patients with chronic pain who are not depressed obtain pain relief with antidepressants. There is evidence that TCAs potentiate opioid analgesia, so they may be useful adjuncts for the treatment of severe persistent pain such as occurs with malignant tumors. Table 13-2 lists some of the painful conditions that respond to TCAs. TCAs are of particular value in the management of neuropathic pain such as occurs in diabetic neuropathy and postherpetic neuralgia, for which there are few other therapeutic options.

The TCAs that have been shown to relieve pain have significant side effects (Table 13-1; **Chap. 452**). Some of these side effects,

TABLE 13-2 Painful Conditions That Respond to Tricyclic Antidepressants

Postherpetic neuralgia ^a
Diabetic neuropathy ^a
Fibromyalgia ^a
Tension headache ^a
Migraine headache ^a
Rheumatoid arthritis ^{a,b}
Chronic low back pain ^b
Cancer
Central poststroke pain

^aControlled trials demonstrate analgesia. ^bControlled studies indicate benefit but not analgesia.

such as orthostatic hypotension, drowsiness, cardiac conduction delay, memory impairment, constipation, and urinary retention, are particularly problematic in elderly patients, and several are additive to the side effects of opioid analgesics. The selective serotonin reuptake inhibitors such as fluoxetine (Prozac) have fewer and less serious side effects than TCAs, but they are much less effective for relieving pain. It is of interest that venlafaxine (Effexor) and duloxetine (Cymbalta), which are nontricyclic antidepressants that block both serotonin and norepinephrine reuptake, appear to retain most of the pain-relieving effect of TCAs with a side effect profile more like that of the selective serotonin reuptake inhibitors. These drugs may be particularly useful in patients who cannot tolerate the side effects of TCAs.

ANTICONVULSANTS AND ANTIARRHYTHMICS

These drugs are useful primarily for patients with neuropathic pain. Phenytoin (Dilantin) and carbamazepine (Tegretol) were first shown to relieve the pain of trigeminal neuralgia ([Chap. 441](#)). This pain has a characteristic brief, shooting, electric shock-like quality. In fact, anticonvulsants seem to be particularly helpful for pains that have such a lancinating quality. Newer anticonvulsants, the calcium channel alpha-2-delta subunit ligands gabapentin (Neurontin) and pregabalin (Lyrica), are effective for a broad range of neuropathic pains. Furthermore, because of their favorable side effect profile, these newer anticonvulsants are often used as first-line agents.

CANNABINOIDS

These agents are widely used for their analgesic properties, although published evidence suggests that any effects are likely to be modest, with small increases in pain threshold reported and variable reductions in clinical pain intensity. Cannabis more consistently reduces the unpleasantness of the pain experience and, in cancer-related pain, can lessen the nausea and vomiting associated with chemotherapy use. *Marijuana and related compounds are discussed in [Chap. 455](#).*

CHRONIC OPIOID MEDICATION

The long-term use of opioids is accepted for patients with pain due to malignant disease. Although opioid use for chronic pain of nonmalignant origin is controversial, it is clear that, for many patients, opioids are the only option that produces meaningful pain relief. This is understandable because opioids are the most potent and have the broadest range of efficacy of any analgesic medications. Although addiction is rare in patients who first use opioids for pain relief, some degree of tolerance and physical dependence is likely with long-term use. Furthermore, studies suggest that long-term opioid therapy may worsen pain in some individuals, termed *opioid-induced hyperalgesia*. Therefore, before embarking on opioid therapy, other options should be explored, and the limitations and risks of opioids should be explained to the patient. It is also important to point out that some opioid analgesic medications have mixed agonist-antagonist properties (e.g., butorphanol and buprenorphine). From a practical standpoint, this means that they

may worsen pain by inducing an abstinence syndrome in patients who are actively being treated with other opioids and are physically dependent.

With long-term outpatient use of orally administered opioids, it may be desirable to use long-acting compounds such as levorphanol, methadone, extended-release morphine or oxycodone, or transdermal fentanyl (Table 13-1). The pharmacokinetic profiles of these drug preparations enable the maintenance of sustained analgesic blood levels, potentially minimizing side effects such as sedation that are associated with high peak plasma levels, and reducing the likelihood of rebound pain associated with a rapid fall in plasma opioid concentration. Extended-release opioid formulations are approved primarily for patients who are already taking other opioids and should not be used as first-line opioids for pain. Although long-acting opioid preparations may provide superior pain relief in patients with a continuous pattern of ongoing pain, others suffer from intermittent severe episodic pain and experience superior pain control and fewer side effects with the periodic use of short-acting opioid analgesics. Constipation is a virtually universal side effect of opioid use and should be treated expectantly. As noted earlier in the discussion of acute pain treatment, a recent advance for patients is the development of peripherally acting opioid antagonists that can reverse the constipation associated with opioid use without interfering with analgesia.

Soon after the introduction of an extended-release oxycodone formulation (OxyContin) in the late 1990s, a dramatic rise in emergency department visits and deaths associated with oxycodone ingestion appeared. This appears to be due primarily to individuals using a prescription opioid nonmedically. Drug-induced deaths have rapidly risen and are now the second leading cause of death in Americans, just behind motor vehicle fatalities. In 2011, the Office of National Drug Control Policy established a multifaceted approach to address prescription drug abuse, including prescription drug monitoring programs (PDMPs) that allow practitioners to determine if patients are receiving prescriptions from multiple providers and use of law enforcement to eliminate improper prescribing practices. In 2016, the Centers for Disease Control and Prevention (CDC) released the *CDC Guideline for Prescribing Opioids for Chronic Pain*, with recommendations for primary care clinicians who are prescribing opioids for chronic noncancer pain. A modified approach to opioid prescribing was published in 2019 by the Health and Human Services Task Force on chronic pain best medical practices. These guidelines address (1) when to initiate or continue opioids for chronic pain; (2) opioid selection, dosage, duration, follow-up, and discontinuation; and (3) assessing risk and addressing harms of opioid use. The recent increase in scrutiny leaves many practitioners hesitant to prescribe opioid analgesics, other than for brief periods to control pain associated with illness or injury. For now, the choice to begin chronic opioid therapy for a given patient is left to the individual practitioner. Pragmatic guidelines for properly selecting and monitoring patients receiving chronic opioid therapy are shown in [Table 13-3](#); a checklist for primary care clinicians prescribing opioids for noncancer pain is shown in [Table 13-4](#).

TREATMENT OF NEUROPATHIC PAIN

It is important to individualize treatment for patients with neuropathic pain. Several general principles should guide therapy: the first is to move quickly to provide relief, and the second is to minimize drug side effects. For example, in patients with postherpetic neuralgia and significant cutaneous hypersensitivity, topical lidocaine (Lidoderm patches) can provide immediate relief without side effects. The anticonvulsants gabapentin or pregabalin (see above) or antidepressants (nortriptyline, desipramine, duloxetine, or venlafaxine) can be used as first-line drugs for patients with neuropathic pain. Systemically administered antiarrhythmic drugs such as lidocaine and mexiletine are less likely to be effective. Although intravenous infusion of lidocaine can provide analgesia for patients with different types of neuropathic pain, the relief is usually transient,

TABLE 13-3 Guidelines for Selecting and Monitoring Patients Receiving Chronic Opioid Therapy (COT) for the Treatment of Chronic, Noncancer Pain**Patient Selection**

- Conduct a history, physical examination, and appropriate testing, including an assessment of risk of substance abuse, misuse, or addiction.
- Consider a trial of COT if pain is moderate or severe, pain is having an adverse impact on function or quality of life, and potential therapeutic benefits outweigh potential harms.
- A benefit-to-harm evaluation, including a history, physical examination, and appropriate diagnostic testing, should be performed and documented before and on an ongoing basis during COT.

Informed Consent and Use of Management Plans

- Informed consent should be obtained. A continuing discussion with the patient regarding COT should include goals, expectations, potential risks, and alternatives to COT.
- Consider using a written COT management plan to document patient and clinician responsibilities and expectations and assist in patient education.

Initiation and Titration

- Initial treatment with opioids should be considered as a therapeutic trial to determine whether COT is appropriate.
- Opioid selection, initial dosing, and titration should be individualized according to the patient's health status, previous exposure to opioids, attainment of therapeutic goals, and predicted or observed harms.

Monitoring

- Reassess patients on COT periodically and as warranted by changing circumstances. Monitoring should include documentation of pain intensity and level of functioning, assessments of progress toward achieving therapeutic goals, presence of adverse events, and adherence to prescribed therapies.
- In patients on COT who are at high risk or who have engaged in aberrant drug-related behaviors, clinicians should periodically obtain urine drug screens or other information to confirm adherence to the COT plan of care.
- In patients on COT not at high risk and not known to have engaged in aberrant drug-related behaviors, clinicians should consider periodically obtaining urine drug screens or other information to confirm adherence to the COT plan of care.

Source: Adapted with permission from R Chou et al: Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. *J Pain* 10:113, 2009.

typically lasting just hours after the cessation of the infusion. The oral lidocaine congener mexiletine is poorly tolerated, producing frequent gastrointestinal adverse effects. There is no consensus on which class of drug should be used as a first-line treatment for any chronically painful condition. However, because relatively high doses of anticonvulsants are required for pain relief, sedation is not uncommon. Sedation is also a problem with TCAs but is much less of a problem with serotonin/norepinephrine reuptake inhibitors (SNRIs; e.g., venlafaxine and duloxetine). Thus, in the elderly or in patients whose daily activities require high-level mental activity, these drugs should be considered the first line. In contrast, opioid medications should be used as a second- or third-line drug class. Although highly effective for many painful conditions, opioids are sedating, and their effect tends to lessen over time, leading to dose escalation and, occasionally, a worsening of pain. A couple of interesting alternatives to pure opioids are two drugs with mixed opioid and norepinephrine reuptake action: tramadol and tapentadol. Tramadol is a relatively weak opioid but is sometimes effective for pain unresponsive to nonopioid analgesics. Tapentadol is a stronger opioid, but its analgesic action is apparently enhanced by the norepinephrine reuptake blockade. Similarly, drugs of different classes can be used in combination to optimize pain control. Repeated injection of botulinum toxin is an emerging approach that is showing some promise in treating focal neuropathic pain, particularly post-herpetic, trigeminal, and post-traumatic neuralgias.

It is worth emphasizing that many patients, especially those with chronic pain, seek medical attention primarily because they are

TABLE 13-4 Centers for Disease Control and Prevention Checklist for Prescribing Opioids for Chronic Pain**For Primary Care Providers Treating Adults (18+) with Chronic Pain ≥3 months, Excluding Cancer, Palliative, and End-of-Life Care****CHECKLIST****WHEN CONSIDERING LONG-TERM OPIOID THERAPY**

- Set realistic goals for pain and function based on diagnosis (e.g., walk around the block).
- Check that nonopiod therapies tried and optimized.
- Discuss benefits and risks (e.g., addiction, overdose) with patient.
- Evaluate risk of harm or misuse.
 - Discuss risk factors with patient.
 - Check prescription drug monitoring program (PDMP) data.
 - Check urine drug screen.
- Set criteria for stopping or continuing opioids.
- Assess baseline pain and function (e.g., Pain, Enjoyment, General Activity [PEG] scale).
- Schedule initial reassessment within 1–4 weeks.
- Prescribe short-acting opioids using lowest dosage on product labeling; match duration to scheduled reassessment.

IF RENEWING WITHOUT A PATIENT VISIT

- Check that return visit is scheduled ≤3 months from last visit.

WHEN REASSESSING AT A PATIENT VISIT

- Continue opioids only after confirming clinically meaningful improvements in pain and function without significant risks or harm.
- Assess pain and function (e.g., PEG); compare results to baseline.
- Evaluate risk of harm or misuse:
 - Observe patient for signs of oversedation or overdose risk. If yes: Taper dose.
 - Check PDMP.
 - Check for opioid use disorder if indicated (e.g., difficulty controlling use). If yes: Refer for treatment.
- Check that nonopiod therapies optimized. Determine whether to continue, adjust, taper, or stop opioids.
- Calculate opioid dosage morphine milligram equivalent (MME).
 - If ≥50 MME/day total (≥ 50 mg hydrocodone; ≥ 33 mg oxycodone), increase frequency of follow-up; consider offering naloxone.
 - Avoid ≥ 90 MME/day total (≥ 90 mg hydrocodone; ≥ 60 mg oxycodone), or carefully justify; consider specialist referral.
- Schedule reassessment at regular intervals (<3 months).

Source: Centers for Disease Control and Prevention, available at: <https://stacks.cdc.gov/view/cdc/38025>. Accessed May 25, 2017 (Public Domain).

suffering and because only physicians can provide the medications required for pain relief. A primary responsibility of all physicians is to minimize the physical and emotional discomfort of their patients. Familiarity with pain mechanisms and analgesic medications is an important step toward accomplishing this aim.

FURTHER READING

- DE VITA MJ et al: Association of cannabinoid administration with experimental pain in healthy adults a systematic review and meta-analysis. *JAMA Psychiatry* 75:1118, 2018.
- DOWELL D et al: CDC guideline for prescribing opioids for chronic pain—United States, 2016. *JAMA* 315:1624, 2016.
- FINNERUP NB et al: Pharmacotherapy for neuropathic pain in adults: A systematic review and meta-analysis. *Lancet Neurol* 14:162, 2015.
- SUN EC et al: Incidence of and risk factors for chronic opioid use among opioid-naïve patients in the postoperative period. *JAMA Intern Med* 176:1286, 2016.
- U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES: Pain management best practices inter-agency task force report: Updates, gaps, inconsistencies, and recommendations. May 2019. <https://www.hhs.gov/ash/advisory-committees/pain/reports/index.html>.



Chest discomfort is among the most common reasons for which patients present for medical attention at either an emergency department (ED) or an outpatient clinic. The evaluation of nontraumatic chest discomfort is inherently challenging owing to the broad variety of possible causes, a minority of which are life-threatening conditions that should not be missed. It is helpful to frame the initial diagnostic assessment and triage of patients with acute chest discomfort around three categories: (1) myocardial ischemia; (2) other cardiopulmonary causes (myopericardial disease, aortic emergencies, and pulmonary conditions); and (3) noncardiopulmonary causes. Although rapid identification of high-risk conditions is a priority of the initial assessment, strategies that incorporate routine liberal use of testing carry the potential for adverse effects of unnecessary investigations.

EPIDEMILOGY AND NATURAL HISTORY

Chest discomfort is one of the three most common reason for visits to the ED in the United States, resulting in 6 to 7 million emergency visits each year. More than 60% of patients with this presentation are hospitalized for further testing, and most of the remainder undergo additional investigation in the ED. Fewer than 15% of evaluated patients are eventually diagnosed with acute coronary syndrome (ACS), with rates of 10–20% in most series of unselected populations, and a rate as low as 5% in some studies. The most common diagnoses are gastrointestinal causes (Fig. 14-1), and as few as 5% are other life-threatening cardiopulmonary conditions. In a large proportion of patients with transient acute chest discomfort, ACS or another acute cardiopulmonary cause is excluded but the cause is not determined. Therefore, the resources and time devoted to the evaluation of chest discomfort *in the absence of a severe cause* are substantial. Nevertheless, historically, a disconcerting 2–6% of patients with chest discomfort of presumed nonischemic etiology who are discharged from the ED were later deemed to have had a missed myocardial infarction (MI). Patients with a missed diagnosis of MI have a 30-day risk of death that is double that of their counterparts who are hospitalized.

The natural histories of ACS, myocarditis, acute pericardial diseases, pulmonary embolism, and aortic emergencies are discussed in Chaps. 270, 273, 274, 275, 279, and 280, respectively. In a study of more than 350,000 patients with unspecified presumed noncardiopulmonary chest discomfort, the mortality rate 1 year after discharge was <2% and did not differ significantly from age-adjusted mortality in the general

population. The estimated rate of major cardiovascular events through 30 days in patients with acute chest pain who had been stratified as low risk was 2.5% in a large population-based study that excluded patients with ST-segment elevation or definite noncardiac chest pain.

CAUSES OF CHEST DISCOMFORT

The major etiologies of chest discomfort are discussed in this section and summarized in Table 14-1. Additional elements of the history, physical examination, and diagnostic testing that aid in distinguishing these causes are discussed in a later section (see “Approach to the Patient”).

■ MYOCARDIAL ISCHEMIA/INJURY

Myocardial ischemia causing chest discomfort, termed *angina pectoris*, is a primary clinical concern in patients presenting with chest symptoms. Myocardial ischemia is precipitated by an imbalance between myocardial oxygen requirements and myocardial oxygen supply, resulting in insufficient delivery of oxygen to meet the heart's metabolic demands. Myocardial oxygen consumption may be elevated by increases in heart rate, ventricular wall stress, and myocardial contractility, whereas myocardial oxygen supply is determined by coronary blood flow and coronary arterial oxygen content. When myocardial ischemia is sufficiently severe and prolonged in duration (as little as 20 min), irreversible cellular injury occurs, resulting in MI.

Ischemic heart disease is most commonly caused by atherosomatous plaque that obstructs one or more of the epicardial coronary arteries. Stable ischemic heart disease (Chap. 273) usually results from the gradual atherosclerotic narrowing of the coronary arteries. *Stable angina* is characterized by ischemic episodes that are typically precipitated by a superimposed increase in oxygen demand during physical exertion and relieved upon resting. Ischemic heart disease becomes unstable, manifest by ischemia at rest or with an escalating pattern, most commonly when rupture or erosion of one or more atherosclerotic lesions triggers coronary thrombosis. Unstable ischemic heart disease is further classified clinically by the presence or absence of detectable acute myocardial injury and the presence or absence of ST-segment elevation on the patient's electrocardiogram (ECG). When acute coronary atherothrombosis occurs, the intracoronary thrombus may be partially obstructive, generally leading to myocardial ischemia in the absence of ST-segment elevation. Unstable ischemic heart disease is classified as *unstable angina* when there is no detectable acute myocardial injury and as *non-ST elevation MI (NSTEMI)* when there is evidence of acute myocardial necrosis (Chap. 274). When the coronary thrombus is acutely and completely occlusive, transmural myocardial ischemia usually ensues, with ST-segment elevation on the ECG and myocardial necrosis leading to a diagnosis of *ST elevation MI (STEMI)*; see Chap. 275).

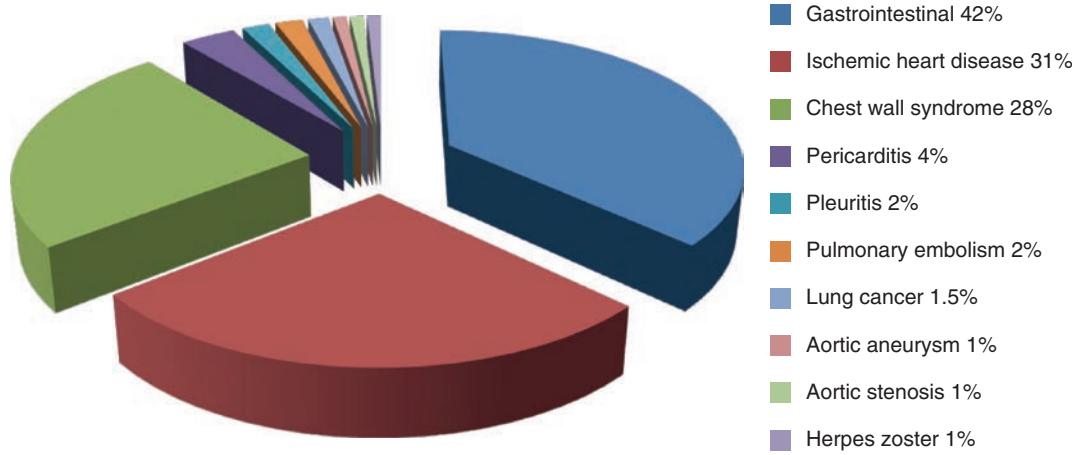


FIGURE 14-1 Distribution of final discharge diagnoses in patients with nontraumatic acute chest pain. (Figure prepared from data in P Fruergaard et al: Eur Heart J 17:1028, 1996.)

TABLE 14-1 Typical Clinical Features of Major Causes of Acute Chest Discomfort

SYSTEM	CONDITION	ONSET/DURATION	QUALITY	LOCATION	ASSOCIATED FEATURES
Cardiopulmonary					
Cardiac	Myocardial ischemia	<i>Stable angina:</i> Precipitated by exertion, cold, or stress; 2–10 min <i>Unstable angina:</i> Increasing pattern or at rest <i>Myocardial infarction:</i> Usually >30 min	Pressure, tightness, squeezing, heaviness, burning	Retrosternal; often radiation to neck, jaw, shoulders, or arms; sometimes epigastric	S_4 gallop or mitral regurgitation murmur (rare) during pain; S_3 or rales if severe ischemia or complication of myocardial infarction
		Pericarditis	Pleuritic, sharp	Retrosternal or toward cardiac apex; may radiate to left shoulder	May be relieved by sitting up and leaning forward; pericardial friction rub
Vascular					
Vascular	Acute aortic syndrome	Sudden onset of unrelenting pain	Tearing or ripping; knifelike	Anterior chest, often radiating to back, between shoulder blades	Associated with hypertension and/or underlying connective tissue disorder; murmur of aortic insufficiency; loss of peripheral pulses
	Pulmonary embolism	Sudden onset	Pleuritic; may manifest as heaviness with massive pulmonary embolism	Often lateral, on the side of the embolism	Dyspnea, tachypnea, tachycardia, and hypotension
	Pulmonary hypertension	Variable; often exertional	Pressure	Substernal	Dyspnea, signs of increased venous pressure
Pulmonary					
Pulmonary	Pneumonia or pleuritis	Variable	Pleuritic	Unilateral, often localized	Dyspnea, cough, fever, rales, occasional rub
	Spontaneous pneumothorax	Sudden onset	Pleuritic	Lateral to side of pneumothorax	Dyspnea, decreased breath sounds on side of pneumothorax
Noncardiopulmonary					
Gastrointestinal					
Gastrointestinal	Esophageal reflux	10–60 min	Burning	Substernal, epigastric	Worsened by postprandial recumbency; relieved by antacids
	Esophageal spasm	2–30 min	Pressure, tightness, burning	Retrosternal	Can closely mimic angina
	Peptic ulcer	Prolonged; 60–90 min after meals	Burning	Epigastric, substernal	Relieved with food or antacids
	Gallbladder disease	Prolonged	Aching or colicky	Epigastric, right upper quadrant; sometimes to the back	May follow meal
Neuromuscular					
Neuromuscular	Costochondritis	Variable	Aching	Sternal	Sometimes swollen, tender, warm over joint; may be reproduced by localized pressure on examination
	Cervical disk disease	Variable; may be sudden	Aching; may include numbness	Arms and shoulders	May be exacerbated by movement of neck
	Trauma or strain	Usually constant	Aching	Localized to area of strain	Reproduced by movement or palpation
	Herpes zoster	Usually prolonged	Sharp or burning	Dermatomal distribution	Vesicular rash in area of discomfort
Psychological	Emotional and psychiatric conditions	Variable; may be fleeting or prolonged	Variable; often manifests as tightness and dyspnea with feeling of panic or doom	Variable; may be retrosternal	Situational factors may precipitate symptoms; history of panic attacks, depression

Clinicians should be aware that unstable ischemic symptoms may also occur predominantly because of increased myocardial oxygen demand (e.g., during intense psychological stress or fever) or because of decreased oxygen delivery due to anemia, hypoxia, or hypotension. However, the term *acute coronary syndrome*, which encompasses unstable angina, NSTEMI, and STEMI, is in general reserved for ischemia precipitated by acute coronary atherothrombosis. In order to guide therapeutic strategies, a standardized system for classification of MI has been expanded to discriminate MI resulting from acute coronary thrombosis (type 1 MI) from MI occurring secondary to other imbalances of myocardial oxygen supply and demand (type 2 MI; see Chap. 274). These conditions are additionally distinguished from nonischemic causes of acute myocardial injury, such as myocarditis.

Other contributors to stable and unstable ischemic heart disease, such as endothelial dysfunction, microvascular disease, and vasoconstriction, may exist alone or in combination with coronary atherosclerosis and may be the dominant cause of myocardial ischemia in some patients. Moreover, nonatherosclerotic processes, including congenital abnormalities of the coronary vessels, myocardial bridging, coronary arteritis, and radiation-induced coronary disease, can lead to coronary obstruction. In addition, conditions associated with extreme myocardial oxygen demand and impaired endocardial blood flow, such as aortic valve disease (Chap. 280), hypertrophic cardiomyopathy, or idiopathic dilated cardiomyopathy (Chap. 259), can precipitate myocardial ischemia in patients with or without underlying obstructive atherosclerosis.

Characteristics of Ischemic Chest Discomfort The clinical characteristics of angina pectoris, often referred to simply as “angina,” are highly similar whether the ischemic discomfort is a manifestation of stable ischemic heart disease, unstable angina, or MI; the exceptions are differences in the pattern and duration of symptoms associated with these syndromes (Table 14-1). Heberden initially described angina as a sense of “strangling and anxiety.” Chest discomfort characteristic of myocardial ischemia is typically described as aching, heavy, squeezing, crushing, or constricting. However, in a substantial minority of patients, the quality of discomfort is extremely vague and may be described as a mild tightness, or merely an uncomfortable feeling, that sometimes is experienced as numbness or a burning sensation. The site of the discomfort is usually retrosternal, but radiation is common and generally occurs down the ulnar surface of the left arm; the right arm, both arms, neck, jaw, or shoulders may also be involved. These and other characteristics of ischemic chest discomfort pertinent to discrimination from other causes of chest pain are discussed later in this chapter (see “Approach to the Patient”).

Stable angina usually begins gradually and reaches its maximal intensity over a period of minutes before dissipating within several minutes with rest or with nitroglycerin. The discomfort typically occurs predictably at a characteristic level of exertion or psychological stress. By definition, unstable angina is manifest by anginal chest discomfort that occurs with progressively lower intensity of physical activity or even at rest. Chest discomfort associated with MI is commonly more severe, is prolonged (usually lasting ≥ 30 min), and is not relieved by rest.

Mechanisms of Cardiac Pain The neural pathways involved in ischemic cardiac pain are poorly understood. Ischemic episodes are thought to excite local chemosensitive and mechanoreceptive receptors that, in turn, stimulate release of adenosine, bradykinin, and other substances that activate the sensory ends of sympathetic and vagal afferent fibers. The afferent fibers traverse the nerves that connect to the upper five thoracic sympathetic ganglia and upper five distal thoracic roots of the spinal cord. From there, impulses are transmitted to the thalamus. Within the spinal cord, cardiac sympathetic afferent impulses may converge with impulses from somatic thoracic structures, and this convergence may be the basis for referred cardiac pain. In addition, cardiac vagal afferent fibers synapse in the nucleus tractus solitarius of the medulla and then descend to the upper cervical spinothalamic tract, and this route may contribute to anginal pain experienced in the neck and jaw.

■ OTHER CARDIOPULMONARY CAUSES

Pericardial and Other Myocardial Diseases (See also Chap. 270)

Inflammation of the pericardium due to infectious or noninfectious causes can be responsible for acute or chronic chest discomfort. The visceral surface and most of the parietal surface of the pericardium are insensitive to pain. Therefore, the pain of pericarditis is thought to arise principally from associated pleural inflammation. Because of this pleural association, the discomfort of pericarditis is usually pleuritic pain that is exacerbated by breathing, coughing, or changes in position. Moreover, owing to the overlapping sensory supply of the central diaphragm via the phrenic nerve with somatic sensory fibers originating in the third to fifth cervical segments, the pain of pleural and pericardial inflammation is often referred to the shoulder and neck. Involvement of the pleural surface of the lateral diaphragm can lead to pain in the upper abdomen.

Acute inflammatory and other nonischemic myocardial diseases can also produce chest discomfort. The symptoms of acute myocarditis are highly varied. Chest discomfort may either originate with inflammatory injury of the myocardium or be due to severe increases in wall stress related to poor ventricular performance. The symptoms of *Takotsubo (stress-related) cardiomyopathy* often start abruptly with chest pain and shortness of breath. This form of cardiomyopathy, in its most recognizable form, is triggered by an emotionally or physically stressful event and may mimic acute MI because of its commonly

associated ECG abnormalities, including ST-segment elevation, and elevated biomarkers of myocardial injury. Observational studies support a predilection for women >50 years of age.

Diseases of the Aorta (See also Chap. 280) Acute aortic dissection (Fig. 14-1) is a less common cause of chest discomfort but is important because of the catastrophic natural history of certain subsets of cases when recognized late or left untreated. Acute aortic syndromes encompass a spectrum of acute aortic diseases related to disruption of the media of the aortic wall. *Aortic dissection* involves a tear in the aortic intima, resulting in separation of the media and creation of a separate “false” lumen. A *penetrating ulcer* has been described as ulceration of an aortic atheromatous plaque that extends through the intima and into the aortic media, with the potential to initiate an intramedial dissection or rupture into the adventitia. *Intramural hematoma* is an aortic wall hematoma with no demonstrable intimal flap, no radiologically apparent intimal tear, and no false lumen. Intramural hematoma can occur due to either rupture of the vasa vasorum or, less commonly, a penetrating ulcer.

Each of these subtypes of acute aortic syndrome typically presents with chest discomfort that is often severe, sudden in onset, and sometimes described as “tearing” in quality. Acute aortic syndromes involving the *ascending* aorta tend to cause pain in the midline of the anterior chest, whereas *descending* aortic syndromes most often present with pain in the back. Therefore, dissections that begin in the ascending aorta and extend to the descending aorta tend to cause pain in the front of the chest that extends toward the back, between the shoulder blades. Proximal aortic dissections that involve the ascending aorta (type A in the Stanford nomenclature) are at high risk for major complications that may influence the clinical presentation, including (1) compromise of the aortic ostia of the coronary arteries, resulting in MI; (2) disruption of the aortic valve, causing acute aortic insufficiency; and (3) rupture of the hematoma into the pericardial space, leading to pericardial tamponade.

Knowledge of the epidemiology of acute aortic syndromes can be helpful in maintaining awareness of this relatively uncommon group of disorders (estimated annual incidence, 3 cases per 100,000 population). Nontraumatic aortic dissections are very rare in the absence of hypertension or conditions associated with deterioration of the elastic or muscular components of the aortic media, including pregnancy, bicuspid aortic disease, or inherited connective tissue diseases, such as Marfan and Ehlers-Danlos syndromes.

Although aortic aneurysms are most often asymptomatic, thoracic aortic aneurysms can cause chest pain and other symptoms by compressing adjacent structures. This pain tends to be steady, deep, and occasionally severe. Aortitis, whether of noninfectious or infectious etiology, in the absence of aortic dissection is a rare cause of chest or back discomfort.

Pulmonary Conditions Pulmonary and pulmonary-vascular conditions that cause chest discomfort usually do so in conjunction with dyspnea and often produce symptoms that have a pleuritic nature.

PULMONARY EMBOLISM (SEE ALSO CHAP. 279) Pulmonary emboli (annual incidence, ~1 per 1000) can produce dyspnea and chest discomfort that is sudden in onset. Typically pleuritic in pattern, the chest discomfort associated with pulmonary embolism may result from (1) involvement of the pleural surface of the lung adjacent to a resultant pulmonary infarction; (2) distention of the pulmonary artery; or (3) possibly, right ventricular wall stress and/or subendocardial ischemia related to acute pulmonary hypertension. The pain associated with small pulmonary emboli is often lateral and pleuritic and is believed to be related to the first of these three possible mechanisms. In contrast, massive pulmonary emboli may cause severe substernal pain that may mimic an MI and that is plausibly attributed to the second and third of these potential mechanisms. Massive or submassive pulmonary embolism may also be associated with syncope, hypotension, and signs of right heart failure. Other typical characteristics that aid in the recognition of pulmonary embolism are discussed later in this chapter (see “Approach to the Patient”).

PNEUMOTHORAX (SEE ALSO CHAP. 294) Primary spontaneous pneumothorax is a rare cause of chest discomfort, with an estimated annual incidence in the United States of 7 per 100,000 among men and <2 per 100,000 among women. Risk factors include male sex, smoking, family history, and Marfan syndrome. The symptoms are usually sudden in onset, and dyspnea may be mild; thus, presentation to medical attention is sometimes delayed. Secondary spontaneous pneumothorax may occur in patients with underlying lung disorders, such as chronic obstructive pulmonary disease, asthma, or cystic fibrosis, and usually produces symptoms that are more severe. Tension pneumothorax is a medical emergency caused by trapped intrathoracic air that precipitates hemodynamic collapse.

Other Pulmonary Parenchymal, Pleural, or Vascular Disease (See also Chaps. 283, 284, and 294) Most pulmonary diseases that produce chest pain, including pneumonia and malignancy, do so because of involvement of the pleura or surrounding structures. Pleurisy is typically described as a knifelike pain that is worsened by inspiration or coughing. In contrast, chronic pulmonary hypertension can manifest as chest pain that may be very similar to angina in its characteristics, suggesting right ventricular myocardial ischemia in some cases. Reactive airways diseases similarly can cause chest tightness associated with breathlessness rather than pleurisy.

■ NONCARDIOPULMONARY CAUSES

Gastrointestinal Conditions (See also Chap. 321) Gastrointestinal disorders are the most common cause of nontraumatic chest discomfort and often produce symptoms that are difficult to discern from more serious causes of chest pain, including myocardial ischemia. Esophageal disorders, in particular, may simulate angina in the character and location of the pain. Gastroesophageal reflux and disorders of esophageal motility are common and should be considered in the differential diagnosis of chest pain (Fig. 14-1 and Table 14-1). The pain of esophageal spasm is commonly an intense, squeezing discomfort that is retrosternal in location and, like angina, may be relieved by nitroglycerin or dihydropyridine calcium channel antagonists. Chest pain can also result from injury to the esophagus, such as a Mallory-Weiss tear or even an esophageal rupture (Boerhaave's syndrome) caused by severe vomiting. Peptic ulcer disease is most commonly epigastric in location but can radiate into the chest (Table 14-1).

Hepatobiliary disorders, including cholecystitis and biliary colic, may mimic acute cardiopulmonary diseases. Although the pain arising from these disorders usually localizes to the right upper quadrant of the abdomen, it is variable and may be felt in the epigastrum and radiate to the back and lower chest. This discomfort is sometimes referred to the scapula or may in rare cases be felt in the shoulder, suggesting diaphragmatic irritation. The pain is steady, usually lasts several hours, and subsides spontaneously, without symptoms between attacks. Pain resulting from pancreatitis is typically aching epigastric pain that radiates to the back.

Musculoskeletal and Other Causes (See also Chap. 360)

Chest discomfort can be produced by any musculoskeletal disorder involving the chest wall or the nerves of the chest wall, neck, or upper limbs. Costochondritis causing tenderness of the costochondral junctions (*Tietze's syndrome*) is relatively common. Cervical radiculitis may manifest as a prolonged or constant aching discomfort in the upper chest and limbs. The pain may be exacerbated by motion of the neck. Occasionally, chest pain can be caused by compression of the brachial plexus by the cervical ribs, and tendinitis or bursitis involving the left shoulder may mimic the radiation of angina. Pain in a dermatomal distribution can also be caused by cramping of intercostal muscles or by herpes zoster (Chap. 193).

Emotional and Psychiatric Conditions As many as 10% of patients who present to EDs with acute chest discomfort have a panic disorder or related condition (Table 14-1). The symptoms may include chest tightness or aching that is associated with a sense of anxiety and difficulty breathing. The symptoms may be prolonged or fleeting.

APPROACH TO THE PATIENT

Chest Discomfort

Given the broad set of potential causes and the heterogeneous risk of serious complications in patients who present with acute nontraumatic chest discomfort, the priorities of the initial clinical encounter include assessment of (1) the patient's clinical stability and (2) the probability that the patient has an underlying cause of the discomfort that may be life-threatening. The high-risk conditions of principal concern are acute cardiopulmonary processes, including ACS, acute aortic syndrome, pulmonary embolism, tension pneumothorax, and pericarditis with tamponade. Fulminant myocarditis also carries a poor prognosis but is usually also manifest by heart failure symptoms. Among noncardiopulmonary causes of chest pain, esophageal rupture likely holds the greatest urgency for diagnosis. Patients with these conditions may deteriorate rapidly despite initially appearing well. The remaining population with noncardiopulmonary conditions has a more favorable prognosis during completion of the diagnostic workup. A rapid targeted assessment for a serious cardiopulmonary cause is of particular relevance for patients with acute ongoing pain who have presented for emergency evaluation. Among patients presenting in the outpatient setting with chronic pain or pain that has resolved, a general diagnostic assessment is reasonably undertaken (see "Outpatient Evaluation of Chest Discomfort," below). A series of questions that can be used to structure the clinical evaluation of patients with chest discomfort is shown in Table 14-2.

HISTORY

The evaluation of nontraumatic chest discomfort relies heavily on the clinical history and physical examination to direct subsequent diagnostic testing. The evaluating clinician should assess the quality, location (including radiation), and pattern (including onset and duration) of the pain as well as any provoking or alleviating factors. The presence of associated symptoms may also be useful in establishing a diagnosis.

Quality of Pain The quality of chest discomfort alone is never sufficient to establish a diagnosis. However, the characteristics of the pain are pivotal in formulating an initial clinical impression and assessing the likelihood of a serious cardiopulmonary process

TABLE 14-2 Considerations in the Assessment of the Patient with Chest Discomfort

1. Could the chest discomfort be due to an acute, potentially life-threatening condition that warrants urgent evaluation and management?			
Unstable ischemic heart disease	Aortic dissection	Pneumothorax	Pulmonary embolism
2. If not, could the discomfort be due to a chronic condition likely to lead to serious complications?			
Stable angina	Aortic stenosis	Pulmonary hypertension	
3. If not, could the discomfort be due to an acute condition that warrants specific treatment?			
Pericarditis	Pneumonia/pleuritis	Herpes zoster	
4. If not, could the discomfort be due to another treatable chronic condition?			
Esophageal reflux		Cervical disk disease	
Esophageal spasm		Arthritis of the shoulder or spine	
Peptic ulcer disease		Costochondritis	
Gallbladder disease		Other musculoskeletal disorders	
Other gastrointestinal conditions		Anxiety state	

Source: Developed by Dr. Thomas H. Lee for the 18th edition of *Harrison's Principles of Internal Medicine*.

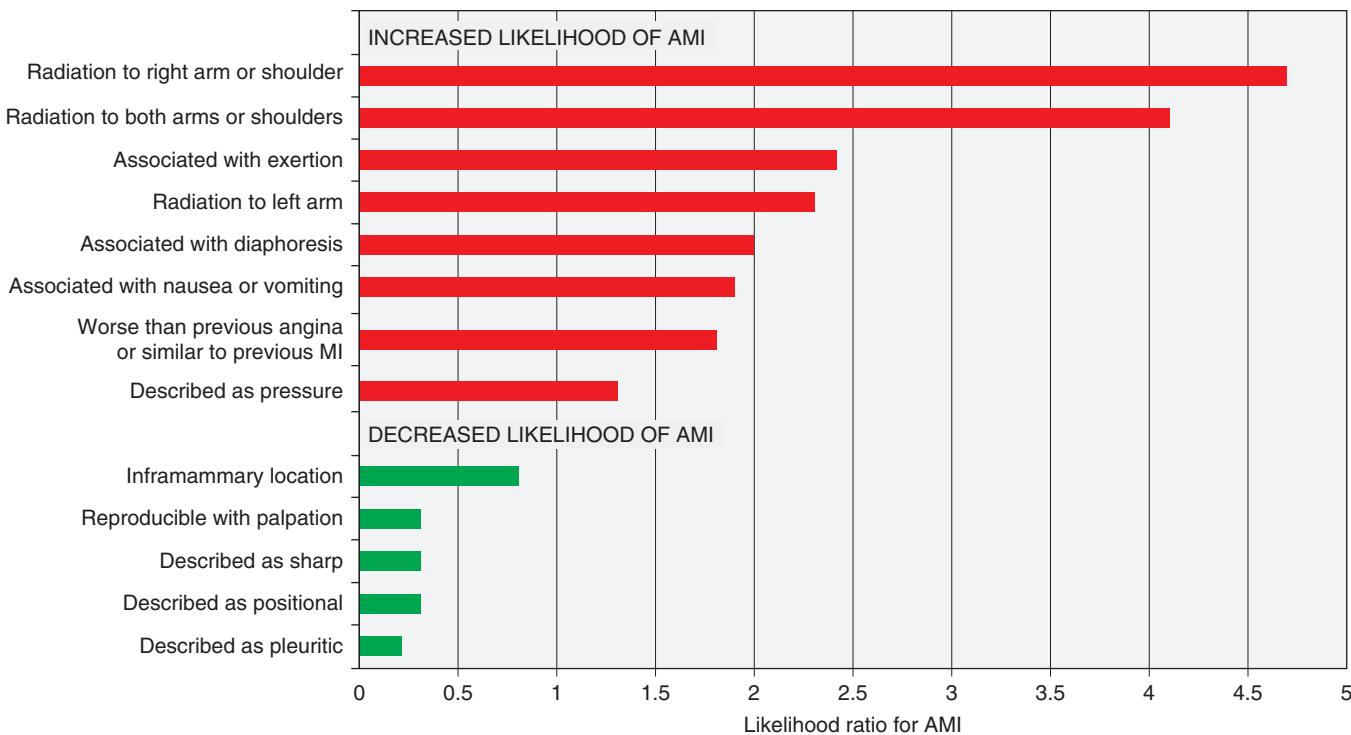


FIGURE 14-2 Association of chest pain characteristics with the probability of acute myocardial infarction (AMI). Note that a subsequent larger study showed a nonsignificant association with radiation to the right arm. (Figure prepared from data in CJ Swap, JT Nagurney: JAMA 294:2623, 2005.)

(Table 14-1), including ACS in particular (Fig. 14-2). Pressure or tightness is consistent with a typical presentation of myocardial ischemic pain. Nevertheless, the clinician must remember that some patients with ischemic chest symptoms deny any “pain” but rather complain of dyspnea or a vague sense of anxiety. The severity of the discomfort has poor diagnostic accuracy. It is often helpful to ask about the similarity of the discomfort to previous definite ischemic symptoms. It is unusual for angina to be sharp, as in knifelike, stabbing, or pleuritic; however, patients sometimes use the word “sharp” to convey the intensity of discomfort rather than the quality. Pleuritic discomfort is suggestive of a process involving the pleura, including pericarditis, pulmonary embolism, or pulmonary parenchymal processes. Less frequently, the pain of pericarditis or massive pulmonary embolism is a steady severe pressure or aching that can be difficult to discriminate from myocardial ischemia. “Tearing” or “ripping” pain is often described by patients with acute aortic dissection. However, acute aortic emergencies also present commonly with knifelike pain. A burning quality can suggest acid reflux or peptic ulcer disease but may also occur with myocardial ischemia. Esophageal pain, particularly with spasm, can be a severe squeezing discomfort identical to angina.

Location of Discomfort A substernal location with radiation to the neck, jaw, shoulder, or arms is typical of myocardial ischemic discomfort. Radiation to both arms has a particularly high association with MI as the etiology. Some patients present with aching in sites of radiated pain as their only symptoms of ischemia. However, pain that is highly localized—e.g., that which can be demarcated by the tip of one finger—is highly unusual for angina. A retrosternal location should prompt consideration of esophageal pain; however, other gastrointestinal conditions usually present with pain that is most intense in the abdomen or epigastrium, with possible radiation into the chest. Angina may also occur in an epigastric location. Pain that occurs solely above the mandible or below the epigastrium is rarely angina. Severe pain radiating to the back, particularly between the shoulder blades, should prompt consideration of an acute aortic syndrome. Radiation to the trapezius ridge is characteristic of pericardial pain and does not usually occur with angina.

Pattern Myocardial ischemic discomfort usually builds over minutes and is exacerbated by activity and mitigated by rest. In contrast, pain that reaches its peak intensity immediately is more suggestive of aortic dissection, pulmonary embolism, or spontaneous pneumothorax. Pain that is fleeting (lasting only a few seconds) is rarely ischemic in origin. Similarly, pain that is constant in intensity for a prolonged period (many hours to days) is unlikely to represent myocardial ischemia if it occurs in the absence of other clinical consequences, such as abnormalities of the ECG, elevation of cardiac biomarkers, or clinical sequelae (e.g., heart failure or hypotension). Both myocardial ischemia and acid reflux may have their onset in the morning.

Provoking and Alleviating Factors Patients with myocardial ischemic pain usually prefer to rest, sit, or stop walking. However, clinicians should be aware of the phenomenon of “warm-up angina” in which some patients experience relief of angina as they continue at the same or even a greater level of exertion (Chap. 273). Alterations in the intensity of pain with changes in position or movement of the upper extremities and neck are less likely with myocardial ischemia and suggest a musculoskeletal etiology. The pain of pericarditis, however, often is worse in the supine position and relieved by sitting upright and leaning forward. Gastroesophageal reflux may be exacerbated by alcohol, some foods, or a reclined position. Relief can occur with sitting.

Exacerbation by eating suggests a gastrointestinal etiology such as peptic ulcer disease, cholecystitis, or pancreatitis. Peptic ulcer disease tends to become symptomatic 60–90 min after meals. However, in the setting of severe coronary atherosclerosis, redistribution of blood flow to the splanchnic vasculature after eating can trigger postprandial angina. The discomfort of acid reflux and peptic ulcer disease is usually diminished promptly by acid-reducing therapies. In contrast with its impact in some patients with angina, physical exertion is very unlikely to alter symptoms from gastrointestinal causes of chest pain. Relief of chest discomfort within minutes after administration of nitroglycerin is suggestive of but not sufficiently sensitive or specific for a definitive diagnosis of myocardial ischemia. Esophageal spasm may also be relieved promptly with

nitroglycerin. A delay of >10 min before relief is obtained after nitroglycerin suggests that the symptoms either are not caused by ischemia or are caused by severe ischemia, such as during acute MI.

Associated Symptoms Symptoms that accompany myocardial ischemia may include diaphoresis, dyspnea, nausea, fatigue, faintness, and eructations. In addition, these symptoms may exist in isolation as anginal equivalents (i.e., symptoms of myocardial ischemia other than typical angina), particularly in women and the elderly. Dyspnea may occur with multiple conditions considered in the differential diagnosis of chest pain and thus is not discriminative, but the presence of dyspnea is important because it suggests a cardiopulmonary etiology. Sudden onset of significant respiratory distress should lead to consideration of pulmonary embolism and spontaneous pneumothorax. Hemoptysis may occur with pulmonary embolism or as blood-tinged frothy sputum in severe heart failure but usually points toward a pulmonary parenchymal etiology of chest symptoms. Presentation with syncope or presyncope should prompt consideration of hemodynamically significant pulmonary embolism or aortic dissection as well as ischemic arrhythmias. Although nausea and vomiting suggest a gastrointestinal disorder, these symptoms may occur in the setting of MI (more commonly inferior MI), presumably because of activation of the vagal reflex or stimulation of left ventricular receptors as part of the Bezold-Jarisch reflex.

Past Medical History The past medical history is useful in assessing the patient for risk factors for coronary atherosclerosis and venous thromboembolism (**Chap. 279**) as well as for conditions that may predispose the patient to specific disorders. For example, a history of connective tissue diseases such as Marfan syndrome should heighten the clinician's suspicion of an acute aortic syndrome or spontaneous pneumothorax. A careful history may elicit clues about depression or prior panic attacks.

PHYSICAL EXAMINATION

In addition to providing an initial assessment of the patient's clinical stability, the physical examination of patients with chest discomfort can provide direct evidence of specific etiologies of chest pain (e.g., unilateral absence of lung sounds) and can identify potential precipitants of acute cardiopulmonary causes of chest pain (e.g., uncontrolled hypertension), relevant comorbid conditions (e.g., obstructive pulmonary disease), and complications of the presenting syndrome (e.g., heart failure). However, because the findings on physical examination may be normal in patients with unstable ischemic heart disease, an unremarkable physical exam is not definitively reassuring.

General The patient's general appearance is helpful in establishing an initial impression of the severity of illness. Patients with acute MI or other acute cardiopulmonary disorders often appear anxious, uncomfortable, pale, cyanotic, or diaphoretic. Patients who are massaging or clutching their chests may describe their pain with a clenched fist held against the sternum (*Levine's sign*). Occasionally, body habitus is helpful—e.g., in patients with Marfan syndrome or the prototypical young, tall, thin man with spontaneous pneumothorax.

Vital Signs Significant tachycardia and hypotension are indicative of important hemodynamic consequences of the underlying cause of chest discomfort and should prompt a rapid survey for the most severe conditions, such as acute MI with cardiogenic shock, massive pulmonary embolism, pericarditis with tamponade, or tension pneumothorax. Acute aortic emergencies usually present with severe hypertension but may be associated with profound hypotension when there is coronary arterial compromise or dissection into the pericardium. Sinus tachycardia is an important manifestation of submassive pulmonary embolism. Tachypnea and hypoxemia point toward a pulmonary cause. The presence of low-grade fever is non-specific because it may occur with MI and with thromboembolism in addition to infection.

Pulmonary Examination of the lungs may localize a primary pulmonary cause of chest discomfort, as in cases of pneumonia, asthma, or pneumothorax. Left ventricular dysfunction from severe ischemia/infarction as well as acute valvular complications of MI or aortic dissection can lead to pulmonary edema, which is an indicator of high risk.

Cardiac The jugular venous pulse is often normal in patients with acute myocardial ischemia but may reveal characteristic patterns with pericardial tamponade or acute right ventricular dysfunction (**Chaps. 239 and 270**). Cardiac auscultation may reveal a third or, more commonly, a fourth heart sound, reflecting myocardial systolic or diastolic dysfunction. Murmurs of mitral regurgitation or a ventricular-septal defect may indicate mechanical complications of STEMI. A murmur of aortic insufficiency may be a complication of ascending aortic dissection. Other murmurs may reveal underlying cardiac disorders contributory to ischemia (e.g., aortic stenosis or hypertrophic cardiomyopathy). Pericardial friction rubs reflect pericardial inflammation.

Abdominal Localizing tenderness on the abdominal exam is useful in identifying a gastrointestinal cause of the presenting syndrome. Abdominal findings are infrequent with purely acute cardiopulmonary problems, except in the case of right-sided heart failure leading to hepatic congestion.

Extremities Vascular pulse deficits may reflect underlying chronic atherosclerosis, which increases the likelihood of coronary artery disease. However, evidence of acute limb ischemia with loss of the pulse and pallor, particularly in the upper extremities, can indicate catastrophic consequences of aortic dissection. Unilateral lower-extremity swelling should raise suspicion about venous thromboembolism.

Musculoskeletal Pain arising from the costochondral and chondrosternal articulations may be associated with localized swelling, redness, or marked localized tenderness. Pain on palpation of these joints is usually well localized and is a useful clinical sign, although deep palpation may elicit pain in the absence of costochondritis. Although palpation of the chest wall often elicits pain in patients with various musculoskeletal conditions, it should be appreciated that chest wall tenderness does not exclude myocardial ischemia. Sensory deficits in the upper extremities may be indicative of cervical disk disease.

ELECTROCARDIOGRAPHY

Electrocardiography is crucial in the evaluation of nontraumatic chest discomfort. The ECG is pivotal for identifying patients with ongoing ischemia as the principal reason for their presentation as well as secondary cardiac complications of other disorders. Professional society guidelines recommend that an ECG be obtained within 10 min of presentation, with the primary goal of identifying patients with ST-segment elevation diagnostic of MI who are candidates for immediate interventions to restore flow in the occluded coronary artery. ST-segment depression and symmetric T-wave inversions at least 0.2 mV in depth are useful for detecting myocardial ischemia in the absence of STEMI and are also indicative of higher risk of death or recurrent ischemia. Serial performance of ECGs (every 30–60 min) is recommended in the ED evaluation of suspected ACS. In addition, an ECG with right-sided lead placement should be considered in patients with clinically suspected ischemia and a nondiagnostic standard 12-lead ECG. Despite the value of the resting ECG, its sensitivity for ischemia is poor—as low as 20% in some studies.

Abnormalities of the ST segment and T wave may occur in a variety of conditions, including pulmonary embolism, ventricular hypertrophy, acute and chronic pericarditis, myocarditis, electrolyte imbalance, and metabolic disorders. Notably, hyperventilation associated with panic disorder can also lead to nonspecific ST and T-wave abnormalities. Pulmonary embolism is most often associated with sinus tachycardia but can also lead to rightward shift of the ECG axis, manifesting as an S-wave in lead I, with a Q-wave

and T-wave in lead III (**Chaps. 240 and 279**). In patients with ST-segment elevation, the presence of diffuse lead involvement not corresponding to a specific coronary anatomic distribution and PR-segment depression can aid in distinguishing pericarditis from acute MI.

CHEST RADIOGRAPHY

(See **Chap. A12**) Plain radiography of the chest is performed routinely when patients present with acute chest discomfort and selectively when individuals who are being evaluated as outpatients have subacute or chronic pain. The chest radiograph is most useful for identifying pulmonary processes, such as pneumonia or pneumothorax. Findings are often unremarkable in patients with ACS, but pulmonary edema may be evident. Other specific findings include widening of the mediastinum in some patients with aortic dissection, Hampton's hump or Westermark's sign in patients with pulmonary embolism (**Chaps. 279 and A12**), or pericardial calcification in chronic pericarditis.

CARDIAC BIOMARKERS

Laboratory testing in patients with acute chest pain is focused on the detection of myocardial injury. Such injury can be detected by the presence of circulating proteins released from damaged cardiomyocytes. Owing to the time necessary for this release, initial biomarkers of injury may be in the normal range, even in patients with STEMI. Cardiac troponin is the preferred biomarker for the diagnosis of MI and should be measured in all patients with suspected ACS. It is not necessary or advisable to measure troponin in patients without suspicion of ACS unless this test is being used specifically for risk stratification (e.g., in pulmonary embolism or heart failure).

The development of cardiac troponin assays with progressively greater analytical sensitivity has facilitated detection of substantially lower blood concentrations of troponin than was previously possible. This evolution permits earlier detection of myocardial injury and more reliable discrimination of changing values, enhances the overall accuracy of a diagnosis of MI, and improves risk stratification in suspected ACS. For these reasons, high-sensitivity assays are generally preferred over prior generation troponin assays. The greater negative predictive value of a negative troponin result with high-sensitivity assays is an advantage in the evaluation of chest pain in the ED. Rapid rule-out protocols that use serial testing and changes in troponin concentration over as short a period as 1–2 h appear to perform well for diagnosis of ACS when using a high-sensitivity troponin assay. Troponin should be measured at presentation and repeated at 1–3 h using high-sensitivity troponin and 3–6 h using conventional troponin assays. Additional troponin measurements may be warranted beyond 3–6 h when the clinical condition still suggests possible ACS or if there is diagnostic uncertainty. In patients presenting more than 2–3 h after symptom onset, a concentration of cardiac troponin, at the time of hospital presentation, below the limit of detection using a high-sensitivity assay may be sufficient to exclude MI with a negative predictive value >99%.

With the use of high-sensitivity assays for troponin, myocardial injury is detected in a larger proportion of patients who have non-ACS cardiopulmonary conditions than with previous, less sensitive assays. Therefore, other aspects of the clinical evaluation are critical to the practitioner's determination of the probability that the symptoms represent ACS. In addition, observation of a change in cardiac troponin concentration between serial samples is necessary for discriminating acute causes of myocardial injury from chronic elevation due to underlying structural heart disease, end-stage renal disease, or the rare presence of interfering antibodies. The diagnosis of MI is reserved for acute myocardial injury that is marked by a rising and/or falling pattern—with at least one value exceeding the 99th percentile reference limit—and that is caused by ischemia. Other nonischemic insults, such as myocarditis, may result in acute myocardial injury but should not be labeled MI (**Fig. 14-3**).

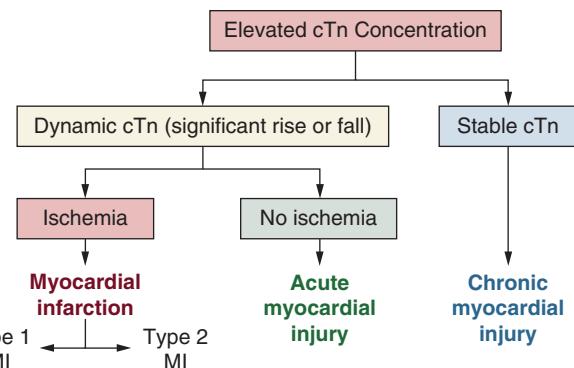


FIGURE 14-3 Clinical classification of patients with elevated cardiac troponin (cTn). MI, myocardial infarction.

Other laboratory assessments may include the D-dimer test to aid in exclusion of pulmonary embolism (**Chap. 279**). Measurement of a B-type natriuretic peptide is useful when considered in conjunction with the clinical history and exam for the diagnosis of heart failure. B-type natriuretic peptides also provide prognostic information among patients with ACS and those with pulmonary embolism.

INTEGRATIVE DECISION-AIDS

Multiple clinical algorithms have been developed to aid in decision-making during the evaluation and disposition of patients with acute nontraumatic chest pain. Such decision-aids estimate either of two closely related but not identical probabilities: (1) the probability of a final diagnosis of ACS and (2) the probability of major cardiac events during short-term follow-up. Such decision-aids are used most commonly to identify patients with a low clinical probability of ACS who are candidates for discharge from the ED, with or without additional noninvasive testing. Goldman and Lee developed one of the first such decision-aids, using only the ECG and risk indicators—hypotension, pulmonary rales, and known ischemic heart disease—to categorize patients into four risk categories ranging from a <1% to a >16% probability of a major cardiovascular complication. Decision-aids used more commonly in current practice are shown in **Fig. 14-4**. Elements common across multiple risk stratification tools are (1) symptoms typical for ACS; (2) older age; (3) risk factors for or known atherosclerosis; (4) ischemic ECG abnormalities; and (5) elevated cardiac troponin level. Although, because of very low specificity, the overall diagnostic performance of such decision-aids is poor (area under the receiver operating curve, 0.55–0.65), in conjunction with the ECG and serial high-sensitivity cardiac troponin, they can help identify patients with a very low probability of ACS (e.g., <1%) or adverse cardiovascular events (<2% at 30 days). Clinical application of such integrated decision-aids or “accelerated diagnostic protocols” has been reported to achieve overall “miss rates” for ACS of <0.5% and may be useful for identifying patients who may be discharged without the need for additional cardiac testing.

Clinicians should differentiate between the algorithms discussed above and risk scores derived for stratification of prognosis (e.g., the TIMI and GRACE risk scores, **Chap. 275**) in patients who already have an established diagnosis of ACS. The latter risk scores were not designed to be used for diagnostic assessment.

CORONARY AND MYOCARDIAL STRESS IMAGING

Among patients for whom other life-threatening causes of chest pain have been reasonably excluded and serial biomarker and clinical assessment have determined the patient to remain eligible for further testing because of intermediate or undetermined risk, diagnostic coronary imaging with coronary computed tomographic (CT) angiography or functional testing, preferably with nuclear or echocardiographic imaging, is recommended. Patient characteristics (e.g., body habitus and renal function), prior cardiac testing,

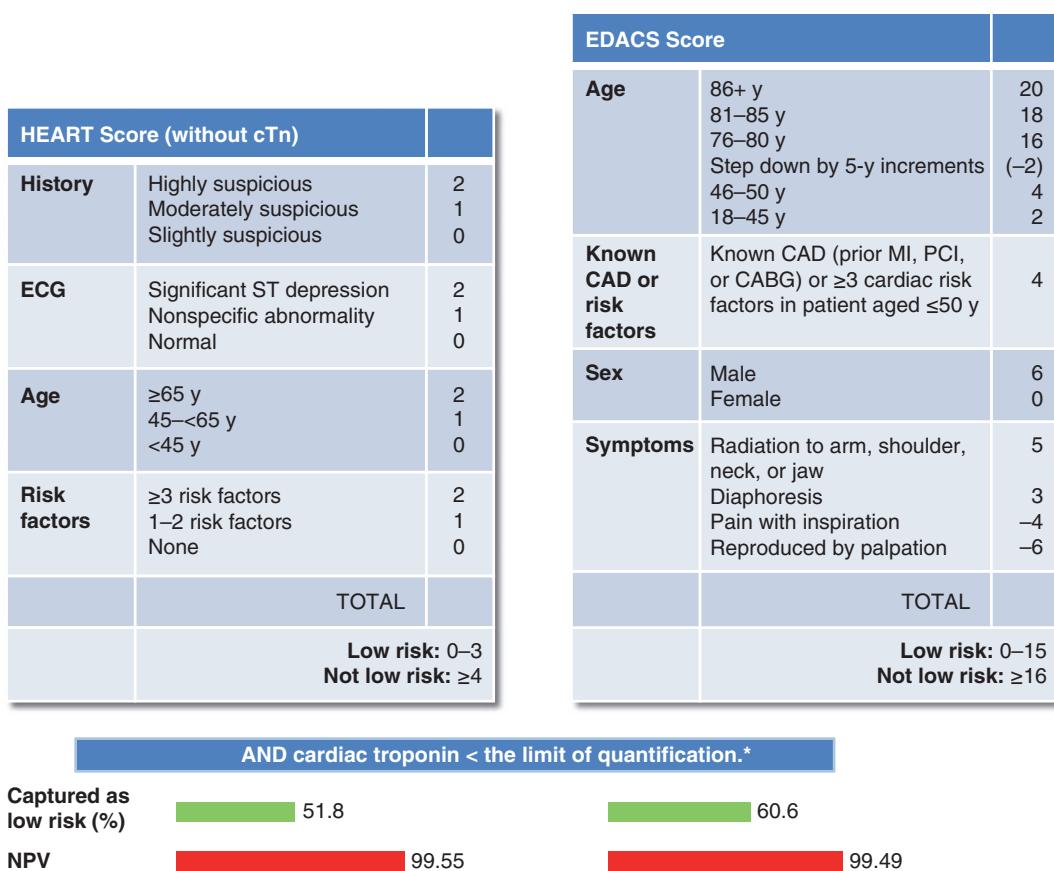


FIGURE 14-4 Examples of decision-aids used in conjunction with serial measurement of cardiac troponin (cTn) for evaluation of acute chest pain. The HEART score was modified by the authors in the presented study and omitting the assignment of 0, 1, or 2 points based on troponin. The negative predictive value (NPV) reported is for the composite endpoint of myocardial infarction (MI), cardiogenic shock, cardiac arrest, and all-cause mortality by 60 days. *Limit of quantification is the lowest analyte concentration that can be quantitatively detected with a total imprecision of ≤20%. CABG, coronary artery bypass graft; CAD, coronary artery disease; ECG, electrocardiogram; PCI, percutaneous coronary intervention. (Figure prepared from data in DG Mark et al: J Am Coll Cardiol 13:606, 2018.)

history of known coronary artery disease, existing contraindications for a given test modality, and patient preferences are considerations when choosing among these diagnostic tests (Chaps. 241 and A9).

CT Angiography (See Chap. 241) CT angiography has emerged as a preferred modality for the evaluation of patients with acute chest discomfort who are candidates for further testing after biomarker and clinical risk assessment. Coronary CT angiography is a sensitive technique for detection of obstructive coronary disease. CT appears to enhance the speed to disposition of patients with a low-intermediate probability for ACS, with its major strength being the negative predictive value of a finding of no significant stenosis or coronary plaque. In addition, contrast-enhanced CT can detect focal areas of myocardial injury in the acute setting. At the same time, CT angiography can exclude aortic dissection, pericardial effusion, and pulmonary embolism.

Stress Nuclear Perfusion Imaging or Stress Echocardiography (See Chaps. 241 and A9) Functional testing with stress nuclear perfusion imaging and stress echocardiography are alternatives for the evaluation of patients with acute chest pain who are candidates for further testing and are preferred over coronary CT angiography in patients with known obstructive epicardial disease. The selection of stress test modality may depend on institutional availability and expertise. Stress testing with myocardial imaging, either with nuclear perfusion imaging or echocardiography, offers superior diagnostic performance over exercise ECG. In patients selected for stress myocardial imaging who are able to exercise, exercise stress is preferred over pharmacologic testing. When available, positron emission tomography offers advantages of improved diagnostic

performance and fewer nondiagnostic studies than single-photon emission CT.

Although functional testing is generally contraindicated in patients with ongoing chest pain, in selected patients with persistent pain and nondiagnostic ECG and biomarker data, resting myocardial perfusion images can be obtained; the absence of any perfusion abnormality substantially reduces the likelihood of coronary artery disease. In such a strategy, used in some centers, those with abnormal rest perfusion imaging, which cannot discriminate between old or new myocardial defects, usually must undergo additional evaluation.

EXERCISE ELECTROCARDIOGRAPHY

Exercise electrocardiography has historically been commonly employed for completion of risk stratification of patients who have undergone an initial evaluation that has not revealed a specific cause of chest discomfort and has identified a low risk of ACS. Early exercise testing is safe in patients without ongoing chest pain or high-risk findings and may assist in refining their prognostic assessment. However, for patients with chest pain for whom both cardiac troponin and clinical risk stratification have determined the patient to have low probability of ACS, there is insufficient evidence that stress testing or cardiac imaging improves their outcomes. This evolution in evidence supports a change from past practice in which outpatient stress testing within 72 hours was broadly used for patients with acute chest pain.

OTHER NONINVASIVE STUDIES

Other noninvasive imaging studies of the chest can be used selectively to provide additional diagnostic and prognostic information on patients with chest discomfort.

Echocardiography Echocardiography (nonstress) is not necessarily routine in patients with chest discomfort. However, in patients with an uncertain diagnosis, particularly those with nondiagnostic ST elevation, ongoing symptoms, or hemodynamic instability, detection of abnormal regional wall motion provides evidence of possible ischemic dysfunction. Echocardiography is diagnostic in patients with mechanical complications of MI or in patients with pericardial tamponade. Transthoracic echocardiography is poorly sensitive for aortic dissection, although an intimal flap may sometimes be detected in the ascending aorta.

MRI (See Chap. 241) Cardiac magnetic resonance (CMR) imaging is an evolving, versatile technique for structural and functional evaluation of the heart and the vasculature of the chest. CMR can be performed as a modality for pharmacologic stress perfusion imaging. Gadolinium-enhanced CMR can provide early detection of MI, defining areas of myocardial necrosis accurately, and can delineate patterns of myocardial disease that are often useful in discriminating ischemic from nonischemic myocardial injury. Although usually not practical for the urgent evaluation of acute chest discomfort, CMR can be a useful modality for cardiac structural evaluation of patients with elevated cardiac troponin levels in the absence of definite coronary artery disease. CMR coronary angiography is in its early stages. MRI also permits highly accurate assessment for aortic dissection but is infrequently used as the first test because CT and transesophageal echocardiography are usually more practical.

■ CRITICAL PATHWAYS FOR ACUTE CHEST DISCOMFORT

Because of the challenges inherent in reliably identifying the small proportion of patients with serious causes of acute chest discomfort while not exposing the larger number of low-risk patients to unnecessary testing and extended ED or hospital evaluations, many medical centers have adopted critical pathways to expedite the assessment and management of patients with nontraumatic chest pain, often in dedicated chest pain units. Such pathways are generally aimed at (1) rapid identification, triage, and treatment of high-risk cardiopulmonary conditions (e.g., STEMI); (2) accurate identification of low-risk patients who can be safely observed in units with less intensive monitoring, undergo early noninvasive testing, or be discharged home; and (3) through more efficient and systematic accelerated diagnostic protocols, safe reduction in costs associated with overuse of testing and unnecessary hospitalizations. In some studies, provision of protocol-driven care in chest pain units has decreased costs and overall duration of hospital evaluation with no detectable excess of adverse clinical outcomes.

■ OUTPATIENT EVALUATION OF CHEST DISCOMFORT

Chest pain is common in outpatient practice, with a lifetime prevalence of 20–40% in the general population. More than 25% of patients with MI have had a related visit with a primary care physician in the previous month. The diagnostic principles are the same as in the ED. However, the pretest probability of an acute cardiopulmonary cause is significantly lower. Therefore, testing paradigms are less intense, with an emphasis on the history, physical examination, and ECG. Moreover, decision-aids developed for settings with a high prevalence of significant cardiopulmonary disease have lower positive predictive value when applied in the practitioner's office. However, in general, if the level of clinical suspicion of ACS is sufficiently high to consider troponin testing, the patient should be referred to the ED for evaluation.

■ FURTHER READING

AMSTERDAM EA et al: Testing of low-risk patients presenting to the emergency department with chest pain: A scientific statement from the American Heart Association. Circulation 122:1756, 2010.
CHAPMAN AR et al: Association of high-sensitivity cardiac troponin I concentration with cardiac outcomes in patients with suspected acute coronary syndrome. JAMA 318:1913, 2017.

FANAROFF AC et al: Does this patient with chest pain have acute coronary syndrome? JAMA 314:1955, 2015.

HSIA RY et al: A national study of the prevalence of life-threatening diagnoses in patients with chest pain. JAMA Intern Med 176:1029, 2016.

MAHLER SA et al: Safely identifying emergency department patients with acute chest pain for early discharge: HEART pathway accelerated diagnostic protocol. Circulation 138:2456, 2018.

15

Abdominal Pain

Danny O. Jacobs



Correctly diagnosing acute abdominal pain can be quite challenging. Few clinical situations require greater judgment, because the most catastrophic of events may be forecast by the subtlest of symptoms and signs. In every instance, the clinician must distinguish those conditions that require urgent intervention from those that do not and can best be managed nonoperatively. A meticulously executed, detailed history and physical examination are critically important for focusing the differential diagnosis and allowing the diagnostic evaluation to proceed expeditiously (**Table 15-1**).

The etiologic classification in **Table 15-2**, although not complete, provides a useful framework for evaluating patients with abdominal pain.

Any patient with abdominal pain of recent onset requires an early and thorough evaluation. The most common causes of abdominal pain on admission are nonspecific abdominal pain, acute appendicitis, pain of urologic origin, and intestinal obstruction. A diagnosis of "acute or surgical abdomen" is not acceptable because of its often misleading and erroneous connotations. Most patients who present with acute abdominal pain will have self-limited disease processes. However, it is important to remember that pain severity does not necessarily correlate with the severity of the underlying condition. And, the presence or absence of various degrees of "hunger" is unreliable as a sole indicator of the severity of intraabdominal disease. The most obvious of "acute abdomens" may not require operative intervention, and the mildest of abdominal pains may herald an urgently correctable disease.

■ SOME MECHANISMS OF PAIN ORIGINATING IN THE ABDOMEN

Inflammation of the Parietal Peritoneum The pain of parietal peritoneal inflammation is steady and aching in character and is located directly over the inflamed area, its exact reference being possible because it is transmitted by somatic nerves supplying the parietal peritoneum. The intensity of the pain is dependent on the type and amount of material to which the peritoneal surfaces are exposed in a given time period. For example, the sudden release of a small quantity

TABLE 15-1 Some Key Components of the Patient's History

Age
Time and mode of onset of the pain
Pain characteristics
Duration of symptoms
Location of pain and sites of radiation
Associated symptoms and their relationship to the pain
Nausea, emesis, and anorexia
Diarrhea, constipation, or other changes in bowel habits
Menstrual history

TABLE 15-2 Some Important Causes of Abdominal Pain

Pain Originating in the Abdomen	
Parietal peritoneal inflammation	Vascular disturbances
Bacterial contamination	Embolism or thrombosis
Perforated appendix or other perforated viscus	Vascular rupture
Pelvic inflammatory disease	Pressure or torsional occlusion
Chemical irritation	Sickle cell anemia
Perforated ulcer	Abdominal wall
Pancreatitis	Distortion or traction of mesentery
Mittelschmerz	Trauma or infection of muscles
Mechanical obstruction of hollow viscera	Distension of visceral surfaces, e.g., by hemorrhage
Obstruction of the small or large intestine	Hepatic or renal capsules
Obstruction of the biliary tree	Inflammation
Obstruction of the ureter	Appendicitis
	Typhoid fever
	Neutropenic enterocolitis or "typhlitis"
Pain Referred from Extraabdominal Source	
Cardiothoracic	Pleurodynia
Acute myocardial infarction	Pneumothorax
Myocarditis, endocarditis, pericarditis	Empyema
Congestive heart failure	Esophageal disease, including spasm, rupture, or inflammation
Pneumonia (especially lower lobes)	Genitalia
Pulmonary embolus	Torsion of the testis
Metabolic Causes	
Diabetes	Acute adrenal insufficiency
Uremia	Familial Mediterranean fever
Hyperlipidemia	Porphyria
Hyperparathyroidism	C1 esterase inhibitor deficiency (angioneurotic edema)
Neurologic/Psychiatric Causes	
Herpes zoster	Spinal cord or nerve root compression
Tabes dorsalis	Functional disorders
Causalgia	Psychiatric disorders
Radiculitis from infection or arthritis	
Toxic Causes	
Lead poisoning	
Insect or animal envenomation	
Black widow spider bites	
Snake bites	
Uncertain Mechanisms	
Narcotic withdrawal	
Heat stroke	

of sterile acidic gastric juice into the peritoneal cavity causes much more pain than the same amount of grossly contaminated neutral feces. Enzymatically active pancreatic juice incites more pain and inflammation than does the same amount of sterile bile containing no potent enzymes. Blood is normally only a mild irritant, and the response to urine is also typically bland, so exposure of blood and urine to the peritoneal cavity may go unnoticed unless it is sudden and massive. Bacterial contamination, such as may occur with pelvic inflammatory disease or perforated distal intestine, causes low-intensity pain until multiplication causes significant amounts of inflammatory mediators to be released. Patients with perforated upper gastrointestinal ulcers may present entirely differently depending on how quickly gastric juices enter the peritoneal cavity and their pH. Thus, the rate at which any inflammatory material irritates the peritoneum is important.

The pain of peritoneal inflammation is invariably accentuated by pressure or changes in tension of the peritoneum, whether produced

by palpation or by movement such as with coughing or sneezing. The patient with peritonitis characteristically lies quietly in bed, preferring to avoid motion, in contrast to the patient with colic, who may be thrashing in discomfort.

Another characteristic feature of peritoneal irritation is tonic reflex spasm of the abdominal musculature, localized to the involved body segment. Its intensity depends on the integrity of the nervous system, the location of the inflammatory process, and the rate at which it develops. Spasm over a perforated retrocecal appendix or perforation into the lesser peritoneal sac may be minimal or absent because of the protective effect of overlying viscera. Catastrophic abdominal emergencies may be associated with minimal or no detectable pain or muscle spasm in obtunded, seriously ill, debilitated, immunosuppressed, or psychotic patients. A slowly developing process also often greatly attenuates the degree of muscle spasm.

Obstruction of Hollow Viscera Intraluminal obstruction classically elicits intermittent or colicky abdominal pain that is not as well localized as the pain of parietal peritoneal irritation. However, the absence of cramping discomfort can be misleading because distention of a hollow viscus may also produce steady pain with only rare paroxysms.

Small-bowel obstruction often presents as poorly localized, intermittent periumbilical or supraumbilical pain. As the intestine progressively dilates and loses muscular tone, the colicky nature of the pain may diminish. With superimposed strangulating obstruction, pain may spread to the lower lumbar region if there is traction on the root of the mesentery. The colicky pain of colonic obstruction is of lesser intensity, is commonly located in the infraumbilical area, and may often radiate to the lumbar region.

Sudden distention of the biliary tree produces a steady rather than colicky type of pain; hence, the term *biliary colic* is misleading. Acute distention of the gallbladder typically causes pain in the right upper quadrant with radiation to the right posterior region of the thorax or to the tip of the right scapula, but discomfort is also not uncommonly found near the midline. Distention of the common bile duct often causes epigastric pain that may radiate to the upper lumbar region. Considerable variation is common, however, so that differentiation between gallbladder or common ductal disease may be impossible.

Gradual dilatation of the biliary tree, as can occur with carcinoma of the head of the pancreas, may cause no pain or only a mild aching sensation in the epigastrium or right upper quadrant. The pain of distention of the pancreatic ducts is similar to that described for distention of the common bile duct but, in addition, is very frequently accentuated by recumbency and relieved by the upright position.

Obstruction of the urinary bladder usually causes dull, low-intensity pain in the suprapubic region. Restlessness, without specific complaint of pain, may be the only sign of a distended bladder in an obtunded patient. In contrast, acute obstruction of the intravesicular portion of the ureter is characterized by severe suprapubic and flank pain that radiates to the penis, scrotum, or inner aspect of the upper thigh. Obstruction of the ureteropelvic junction manifests as pain near the costovertebral angle, whereas obstruction of the remainder of the ureter is associated with flank pain that often extends into the same side of the abdomen.

Vascular Disturbances A frequent misconception is that pain due to intraabdominal vascular disturbances is sudden and catastrophic in nature. Certain disease processes, such as embolism or thrombosis of the superior mesenteric artery or impending rupture of an abdominal aortic aneurysm, can certainly be associated with diffuse, severe pain. Yet, just as frequently, the patient with occlusion of the superior mesenteric artery only has mild continuous or cramping diffuse pain for 2 or 3 days before vascular collapse or findings of peritoneal inflammation appear. The early, seemingly insignificant discomfort is caused by hyperperistalsis rather than peritoneal inflammation. Indeed, absence of tenderness and rigidity in the presence of continuous, diffuse pain (e.g., "pain out of proportion to physical findings") in a patient likely to have vascular disease is quite characteristic of occlusion of the superior mesenteric artery. Abdominal pain with radiation to the sacral region,

flank, or genitalia should always signal the possible presence of a rupturing abdominal aortic aneurysm. This pain may persist over a period of several days before rupture and collapse occur.

Abdominal Wall Pain arising from the abdominal wall is usually constant and aching. Movement, prolonged standing, and pressure accentuate the discomfort and associated muscle spasm. In the relatively rare case of hematoma of the rectus sheath, now most frequently encountered in association with anticoagulant therapy, a mass may be present in the lower quadrants of the abdomen. Simultaneous involvement of muscles in other parts of the body usually serves to differentiate myositis of the abdominal wall from other processes that might cause pain in the same region.

■ REFERRED PAIN IN ABDOMINAL DISEASE

Pain referred to the abdomen from the thorax, spine, or genitalia may present a diagnostic challenge because diseases of the upper part of the abdominal cavity such as acute cholecystitis or perforated ulcer may be associated with intrathoracic complications. A most important, yet often forgotten, dictum is that the possibility of intrathoracic disease must be considered in every patient with abdominal pain, especially if the pain is in the upper abdomen.

Systematic questioning and examination directed toward detecting myocardial or pulmonary infarction, pneumonia, pericarditis, or esophageal disease (the intrathoracic diseases that most often masquerade as abdominal emergencies) will often provide sufficient clues to establish the proper diagnosis. Diaphragmatic pleuritis resulting from pneumonia or pulmonary infarction may cause pain in the right upper quadrant and pain in the supraclavicular area, the latter radiation to be distinguished from the referred subscapular pain caused by acute distention of the extrahepatic biliary tree. The ultimate decision as to the origin of abdominal pain may require deliberate and planned observation over a period of several hours, during which repeated questioning and examination will provide the diagnosis or suggest the appropriate studies.

Referred pain of thoracic origin is often accompanied by splinting of the involved hemithorax with respiratory lag and a decrease in excursion more marked than that seen in the presence of intraabdominal disease. In addition, apparent abdominal muscle spasm caused by referred pain will diminish during the inspiratory phase of respiration, whereas it persists throughout both respiratory phases if it is of abdominal origin. Palpation over the area of referred pain in the abdomen also does not usually accentuate the pain and, in many instances, actually seems to relieve it.

Thoracic disease and abdominal disease frequently coexist and may be difficult or impossible to differentiate. For example, the patient with known biliary tract disease often has epigastric pain during myocardial infarction, or biliary colic may be referred to the precordium or left shoulder in a patient who has suffered previously from angina pectoris. **For an explanation of the radiation of pain to a previously diseased area, see Chap. 13.**

Referred pain from the spine, which usually involves compression or irritation of nerve roots, is characteristically intensified by certain motions such as cough, sneeze, or strain and is associated with hyperesthesia over the involved dermatomes. Pain referred to the abdomen from the testes or seminal vesicles is generally accentuated by the slightest pressure on either of these organs. The abdominal discomfort experienced is of dull, aching character and is poorly localized.

■ METABOLIC ABDOMINAL CRISES

Pain of metabolic origin may simulate almost any other type of intraabdominal disease. Several mechanisms may be at work. In certain instances, such as hyperlipidemia, the metabolic disease itself may be accompanied by an intraabdominal process such as pancreatitis, which can lead to unnecessary laparotomy unless recognized. C1 esterase deficiency associated with angioneurotic edema is often associated with episodes of severe abdominal pain. Whenever the cause of

abdominal pain is obscure, a metabolic origin always must be considered. Abdominal pain is also the hallmark of familial Mediterranean fever (**Chap. 369**).

The pain of porphyria and of lead colic is usually difficult to distinguish from that of intestinal obstruction, because severe hyperperistalsis is a prominent feature of both. The pain of uremia or diabetes is nonspecific, and the pain and tenderness frequently shift in location and intensity. Diabetic acidosis may be precipitated by acute appendicitis or intestinal obstruction, so if prompt resolution of the abdominal pain does not result from correction of the metabolic abnormalities, an underlying organic problem should be suspected. Black widow spider bites produce intense pain and rigidity of the abdominal muscles and back, an area infrequently involved in intraabdominal disease.

■ IMMUNOCOMPROMISE

Evaluating and diagnosing causes of abdominal pain in immunosuppressed or otherwise immunocompromised patients is very difficult. This includes those who have undergone organ transplantation; who are receiving immunosuppressive treatments for autoimmune diseases, chemotherapy, or glucocorticoids; who have AIDS; and who are very old. In these circumstances, normal physiologic responses may be absent or masked. In addition, unusual infections may cause abdominal pain where the etiologic agents include cytomegalovirus, mycobacteria, protozoa, and fungi. These pathogens may affect all gastrointestinal organs, including the gallbladder, liver, and pancreas, as well as the gastrointestinal tract, causing occult or overtly symptomatic perforations of the latter. Splenic abscesses due to *Candida* or *Salmonella* infection should also be considered, especially when evaluating patients with left upper quadrant or left flank pain. Acalculous cholecystitis may be observed in immunocompromised patients or those with AIDS, where it is often associated with cryptosporidiosis or cytomegalovirus infection.

Neutropenic enterocolitis (typhlitis) is often identified as a cause of abdominal pain and fever in some patients with bone marrow suppression due to chemotherapy. Acute graft-versus-host disease should be considered in this circumstance. Optimal management of these patients requires meticulous follow-up including serial examinations to assess the need for more surgical intervention, for example, to address perforation.

■ NEUROGENIC CAUSES

Diseases that injure sensory nerves may cause causalgic pain. This pain has a burning character and is usually limited to the distribution of a given peripheral nerve. Stimuli that are normally not painful such as touch or a change in temperature may be causalgic and are often present even at rest. The demonstration of irregularly spaced cutaneous "pain spots" may be the only indication that an old nerve injury exists. Even though the pain may be precipitated by gentle palpation, rigidity of the abdominal muscles is absent, and the respirations are not usually disturbed. Distention of the abdomen is uncommon, and the pain has no relationship to food intake.

Pain arising from spinal nerves or roots comes and goes suddenly and is of a lancinating type (**Chap. 17**). It may be caused by herpes zoster, impingement by arthritis, tumors, a herniated nucleus pulposus, diabetes, or syphilis. It is not associated with food intake, abdominal distention, or changes in respiration. Severe muscle spasms, when present, may be relieved by, but are certainly not accentuated by, abdominal palpation. The pain is made worse by movement of the spine and is usually confined to a few dermatomes. Hyperesthesia is very common.

Pain due to functional causes conforms to none of the aforementioned patterns. Mechanisms of disease are not clearly established. Irritable bowel syndrome (IBS) is a functional gastrointestinal disorder characterized by abdominal pain and altered bowel habits. The diagnosis is made on the basis of clinical criteria (**Chap. 327**) and after exclusion of demonstrable structural abnormalities. The episodes of abdominal pain may be brought on by stress, and the pain varies considerably in type and location. Nausea and vomiting are rare. Localized

tenderness and muscle spasm are inconsistent or absent. The causes of IBS or related functional disorders are not yet fully understood.

APPROACH TO THE PATIENT

Abdominal Pain

Few abdominal conditions require such urgent operative intervention that an orderly approach needs to be abandoned, no matter how ill the patient is. Only patients with exsanguinating intraabdominal hemorrhage (e.g., ruptured aneurysm) must be rushed to the operating room immediately, but in such instances, only a few minutes are required to assess the critical nature of the problem. Under these circumstances, all obstacles must be swept aside, adequate venous access for fluid replacement obtained, and the operation begun. Unfortunately, many of these patients may die in the radiology department or the emergency room while awaiting unnecessary examinations. *There are no absolute contraindications to operation when massive intraabdominal hemorrhage is present.* Fortunately, this situation is relatively rare. This statement does not necessarily apply to patients with intraluminal gastrointestinal hemorrhage, who can often be managed by other means (Chap. 48). In these patients, obtaining a *detailed history when possible* can be extremely helpful even though it can be laborious and time-consuming. Decision-making regarding next steps is facilitated and a reasonably accurate diagnosis can be made before any further diagnostic testing is undertaken.

In cases of *acute* abdominal pain, a diagnosis can be readily established in most instances, whereas success is not so frequent in patients with *chronic* pain. IBS is one of the most common causes of abdominal pain and must always be kept in mind (Chap. 327). The location of the pain can assist in narrowing the differential diagnosis (Table 15-3); however, the *chronological sequence of events* in the

patient's history is often more important than the pain's location. Careful attention should be paid to the extraabdominal regions. Narcotics or analgesics should *not* be withheld until a definitive diagnosis or a definitive plan has been formulated; obfuscation of the diagnosis by adequate analgesia is unlikely.

An accurate menstrual history in a female patient is essential. It is important to remember that normal anatomic relationships can be significantly altered by the gravid uterus. Abdominal and pelvic pain may occur during pregnancy due to conditions that do not require operation. Lastly, some otherwise noteworthy laboratory values (e.g., leukocytosis) may represent the normal physiologic changes of pregnancy.

In the examination, simple critical inspection of the patient, for example, of facies, position in bed, and respiratory activity, provides valuable clues. The amount of information to be gleaned is directly proportional to the *gentleness* and thoroughness of the examiner. Once a patient with peritoneal inflammation has been examined briskly, accurate assessment by the next examiner becomes almost impossible. Eliciting rebound tenderness by sudden release of a deeply palpating hand in a patient with suspected peritonitis is cruel and unnecessary. The same information can be obtained by gentle percussion of the abdomen (rebound tenderness on a miniature scale), a maneuver that can be far more precise and localizing. Asking the patient to cough will elicit true rebound tenderness without the need for placing a hand on the abdomen. Furthermore, the forceful demonstration of rebound tenderness will startle and induce protective spasm in a nervous or worried patient in whom true rebound tenderness is not present. A palpable gallbladder will be missed if palpation is so aggressive that voluntary muscle spasm becomes superimposed on involuntary muscular rigidity.

As with history taking, sufficient time should be spent in the examination. Abdominal signs may be minimal but, nevertheless, if accompanied by consistent symptoms, may be exceptionally meaningful. Abdominal signs may be virtually or totally absent in cases of pelvic peritonitis, so careful *pelvic and rectal examinations are mandatory in every patient with abdominal pain.* Tenderness on pelvic or rectal examination in the absence of other abdominal signs can be caused by operative indications such as perforated appendicitis, diverticulitis, twisted ovarian cyst, and many others. Much attention has been paid to the presence or absence of peristaltic sounds, their quality, and their frequency. Auscultation of the abdomen is one of the least revealing aspects of the physical examination of a patient with abdominal pain. Catastrophes such as a strangulating small-intestinal obstruction or perforated appendicitis may occur in the presence of normal peristaltic sounds. Conversely, when the proximal part of the intestine above obstruction becomes markedly distended and edematous, peristaltic sounds may lose the characteristics of borborygmi and become weak or absent, even when peritonitis is not present. It is usually the severe chemical peritonitis of sudden onset that is associated with the truly silent abdomen.

Laboratory examinations may be valuable in assessing the patient with abdominal pain, yet, with few exceptions, they rarely establish a diagnosis. Leukocytosis should never be the single deciding factor as to whether or not operation is indicated. A white blood cell count $>20,000/\mu\text{L}$ may be observed with perforation of a viscus, but pancreatitis, acute cholecystitis, pelvic inflammatory disease, and intestinal infarction may also be associated with marked leukocytosis. A normal white blood cell count is not rare in cases of perforation of abdominal viscera. A diagnosis of anemia may be more helpful than the white blood cell count, especially when combined with the history.

The urinalysis may reveal the state of hydration or rule out severe renal disease, diabetes, or urinary infection. Blood urea nitrogen, glucose, and serum bilirubin levels and liver function tests may be

TABLE 15-3 Differential Diagnoses of Abdominal Pain by Location

Right Upper Quadrant	Epigastric	Left Upper Quadrant
Cholecystitis	Peptic ulcer disease	Splenic infarct
Cholangitis	Gastritis	Splenic rupture
Pancreatitis	GERD	Splenic abscess
Pneumonia/empyema	Pancreatitis	Gastritis
Pleurisy/pleurodynia	Myocardial infarction	Gastric ulcer
Subdiaphragmatic abscess	Pericarditis	Pancreatitis
Hepatitis	Ruptured aortic aneurysm	Subdiaphragmatic abscess
Budd-Chiari syndrome	Esophagitis	
Right Lower Quadrant	Periumbilical	Left Lower Quadrant
Appendicitis	Early appendicitis	Diverticulitis
Salpingitis	Gastroenteritis	Salpingitis
Inguinal hernia	Bowel obstruction	Inguinal hernia
Ectopic pregnancy	Ruptured aortic aneurysm	Ectopic pregnancy
Nephrolithiasis		Nephrolithiasis
Inflammatory bowel disease		Irritable bowel syndrome
Mesenteric lymphadenitis		Inflammatory bowel disease
Typhlitis		
Diffuse Nonlocalized Pain		
Gastroenteritis	Malaria	
Mesenteric ischemia	Familial Mediterranean fever	
Bowel obstruction	Metabolic diseases	
Irritable bowel syndrome	Psychiatric disease	
Peritonitis		
Diabetes		

Abbreviation: GERD, gastroesophageal reflux disease.

helpful. Serum amylase levels may be increased by many diseases other than pancreatitis, for example, perforated ulcer, strangulating intestinal obstruction, and acute cholecystitis; thus, elevations of serum amylase do not rule in or rule out the need for an operation.

Plain and upright or lateral decubitus radiographs of the abdomen have limited utility and may be unnecessary in some patients who have substantial evidence of some diseases such as acute appendicitis or strangulated external hernia. Where the indications for surgical or medical intervention are not clear, low-dose computed tomography is preferred to abdominal radiography when evaluating nontraumatic acute abdominal pain.

Very rarely, barium or water-soluble contrast study of the upper part of the gastrointestinal tract is an appropriate radiographic investigation and may demonstrate partial intestinal obstruction that may elude diagnosis by other means. If there is any question of obstruction of the colon, oral administration of barium sulfate should be avoided. On the other hand, in cases of suspected colonic obstruction (without perforation), a contrast enema may be diagnostic.

In the absence of trauma, peritoneal lavage has been replaced as a diagnostic tool by CT scanning and laparoscopy. Ultrasonography has proved to be useful in detecting an enlarged gallbladder or pancreas, the presence of gallstones, an enlarged ovary, or a tubal pregnancy. Laparoscopy is especially helpful in diagnosing pelvic conditions, such as ovarian cysts, tubal pregnancies, salpingitis, acute appendicitis, and other disease processes. Laparoscopy has a particular advantage over imaging in that the underlying etiologic condition can often be definitively addressed.

Radioisotopic hepatobiliary iminodiacetic acid scans (HIDAs) may help differentiate acute cholecystitis or biliary colic from acute pancreatitis. A CT scan may demonstrate an enlarged pancreas, ruptured spleen, or thickened colonic or appendiceal wall and streaking of the mesocolon or mesoappendix characteristic of diverticulitis or appendicitis.

Sometimes, even under the best circumstances with all available aids and with the greatest of clinical skill, a definitive diagnosis cannot be established at the time of the initial examination. And, in some cases, operation may be indicated based on clinical grounds alone. Should that decision be questionable, watchful waiting with repeated questioning and examination will often elucidate the true nature of the illness and indicate the proper course of action.

ACKNOWLEDGMENT

The author gratefully acknowledges the enormous contribution to this chapter and the approach it espouses of William Silen, who wrote this chapter for many editions.

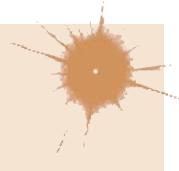
FURTHER READING

- BHANGU A et al: Acute appendicitis: Modern understanding of pathogenesis, diagnosis and management. *Lancet* 386:1278, 2015.
- CARTWRIGHT SL, KNUDSON MP: Diagnostic imaging of acute abdominal pain in adults. *Am Fam Phys* 91:452, 2015.
- HUCKINS DS et al: Diagnostic performance of a biomarker panel as a negative predictor for acute appendicitis in acute emergency department patients with abdominal pain. *Am J Emerg Med* 35:418, 2017.
- NAYOR J et al: Tracing the cause of abdominal pain. *N Engl J Med* 375:e8, 2016.
- PHILLIPS MT: Clinical yield of computed tomography scans in the emergency department for abdominal pain. *J Invest Med* 64:542, 2016.
- SILEN W, COPE Z: *Cope's Early Diagnosis of the Acute Abdomen*, 22nd ed. New York, Oxford University Press, 2010.

16

Headache

Peter J. Goadsby



Headache is among the most common reasons patients seek medical attention and is responsible, on a global basis, for more disability than any other neurologic problem. Diagnosis and management are based on a careful clinical approach augmented by an understanding of the anatomy, physiology, and pharmacology of the nervous system pathways mediating the various headache syndromes. This chapter will focus on the general approach to a patient with headache; migraine and other primary headache disorders are discussed in [Chap. 430](#).

GENERAL PRINCIPLES

A classification system developed by the International Headache Society (www.ihf-headache.org/en/resources/guidelines/) characterizes headache as primary or secondary ([Table 16-1](#)). Primary headaches are those in which headache and its associated features are the disorder itself, whereas secondary headaches are those caused by exogenous disorders (Headache Classification Committee of the International Headache Society, 2018). Primary headache often results in considerable disability and a decrease in the patient's quality of life. Mild secondary headache, such as that seen in association with upper respiratory tract infections, is common but rarely worrisome. Life-threatening headache is relatively uncommon, but vigilance is required in order to recognize and appropriately treat such patients.

ANATOMY AND PHYSIOLOGY OF HEADACHE

Pain usually occurs when peripheral nociceptors are stimulated in response to tissue injury, visceral distension, or other factors ([Chap. 13](#)). In such situations, pain perception is a normal physiologic response mediated by a healthy nervous system. Pain can also result when pain-producing pathways of the peripheral or central nervous system (CNS) are damaged or activated inappropriately. Headache may originate from either or both mechanisms. Relatively few cranial structures are pain producing; these include the scalp, meningeal arteries, dural sinuses, falx cerebri, and proximal segments of the large pial arteries. The ventricular ependyma, choroid plexus, pial veins, and much of the brain parenchyma are not pain producing.

The key structures involved in primary headache are the following:

- The large intracranial vessels and dura mater, and the peripheral terminals of the trigeminal nerve that innervate these structures
- The caudal portion of the trigeminal nucleus, which extends into the dorsal horns of the upper cervical spinal cord and receives input from the first and second cervical nerve roots (the trigeminocervical complex)
- Rostral pain-processing regions, such as the ventroposteromedial thalamus and the cortex
- The pain-modulatory systems in the brain that modulate input from the trigeminal nociceptors at all levels of the pain-processing pathways and influence vegetative functions, such as the hypothalamus and brainstem

TABLE 16-1 Common Causes of Headache

PRIMARY HEADACHE		SECONDARY HEADACHE	
TYPE	%	TYPE	%
Tension-type	69	Systemic infection	63
Migraine	16	Head injury	4
Idiopathic stabbing	2	Vascular disorders	1
Exertional	1	Subarachnoid hemorrhage	<1
Cluster	0.1	Brain tumor	0.1

Source: After J Olesen et al: *The Headaches*. Philadelphia, Lippincott Williams & Wilkins, 2005.

The *trigeminovascular system* innervates the large intracranial vessels and dura mater via the trigeminal nerve. Cranial autonomic symptoms, such as lacrimation, conjunctival injection, nasal congestion, rhinorrhea, periorbital swelling, aural fullness, and ptosis, are prominent in the trigeminal autonomic cephalgias (TACs), including cluster headache and paroxysmal hemicrania, and may also be seen in migraine, even in children. These autonomic symptoms reflect activation of cranial parasympathetic pathways, and functional imaging studies indicate that vascular changes in migraine and cluster headache, when present, are similarly driven by these cranial autonomic systems. Thus, they are secondary, and not causative, events in the headache cascade. Moreover, they can often be mistaken for symptoms or signs of cranial sinus inflammation, which is then overdiagnosed and inappropriately managed. Migraine and other primary headache types are not “vascular headaches”; these disorders do not reliably manifest vascular changes, and treatment outcomes cannot be predicted by vascular effects. Migraine is a brain disorder and is best understood and managed as such.

■ CLINICAL EVALUATION OF ACUTE, NEW-ONSET HEADACHE

The patient who presents with a new, severe headache has a differential diagnosis that is quite different from the patient with recurrent headaches over many years. In new-onset and severe headache, the probability of finding a potentially serious cause is considerably greater than in recurrent headache. Patients with recent onset of pain require prompt evaluation and appropriate treatment. Serious causes to be considered include meningitis, subarachnoid hemorrhage, epidural or subdural hematoma, glaucoma, tumor, and purulent sinusitis. When worrisome symptoms and signs are present (**Table 16-2**), rapid diagnosis and management are critical.

A careful neurologic examination is an essential first step in the evaluation. In most cases, patients with an abnormal examination or a history of recent-onset headache should be evaluated by a computed tomography (CT) or magnetic resonance imaging (MRI) study of the brain. As an initial screening procedure for intracranial pathology in this setting, CT and MRI methods appear to be equally sensitive. In some circumstances, a lumbar puncture (LP) is also required, unless a benign etiology can be otherwise established. A general evaluation of acute headache might include cranial arteries by palpation; cervical spine by the effect of passive movement of the head and by imaging; the investigation of cardiovascular and renal status by blood pressure monitoring and urine examination; and eyes by funduscopy, intraocular pressure measurement, and refraction.

The patient's psychological state should also be evaluated because a relationship exists between head pain, depression, and anxiety. This is intended to identify comorbidity rather than provide an explanation for the headache, because troublesome headache is seldom simply caused by mood change. Although it is notable that medicines with antidepressant actions are also effective in the preventive treatment

of both tension-type headache and migraine, each symptom must be treated optimally.

Underlying recurrent headache disorders may be activated by pain that follows otologic or endodontic surgical procedures. Thus, pain about the head as the result of diseased tissue or trauma may reawaken an otherwise quiescent migraine syndrome. Treatment of the headache is largely ineffective until the cause of the primary problem is addressed.

Serious underlying conditions that are associated with headache are described below. Brain tumor is a rare cause of headache and even less commonly a cause of severe pain. The vast majority of patients presenting with severe headache have a benign cause.

■ SECONDARY HEADACHE

The management of secondary headache focuses on diagnosis and treatment of the underlying condition.

■ MENINGITIS

Acute, severe headache with stiff neck and fever suggests meningitis. LP is mandatory. Often there is striking accentuation of pain with eye movement. Meningitis can be easily mistaken for migraine in that the cardinal symptoms of pounding headache, photophobia, nausea, and vomiting are frequently present, perhaps reflecting the underlying biology of some of the patients.

Meningitis is discussed in Chaps. 138 and 139.

■ INTRACRANIAL HEMORRHAGE

Acute, maximal in <5 min, severe headache lasting >5 min with stiff neck but without fever suggests subarachnoid hemorrhage. A ruptured aneurysm, arteriovenous malformation, or intraparenchymal hemorrhage may also present with headache alone. Rarely, if the hemorrhage is small or below the foramen magnum, the head CT scan can be normal. Therefore, LP may be required to diagnose definitively subarachnoid hemorrhage.

Subarachnoid hemorrhage is discussed in Chap. 429, and intracranial hemorrhage in Chap. 428.

■ BRAIN TUMOR

Approximately 30% of patients with brain tumors consider headache to be their chief complaint. The head pain is usually nondescript—an intermittent deep, dull aching of moderate intensity, which may worsen with exertion or change in position and may be associated with nausea and vomiting. This pattern of symptoms results from migraine far more often than from brain tumor. The headache of brain tumor disturbs sleep in about 10% of patients. Vomiting that precedes the appearance of headache by weeks is highly characteristic of posterior fossa brain tumors. A history of amenorrhea or galactorrhea should lead one to question whether a prolactin-secreting pituitary adenoma (or polycystic ovary syndrome) is the source of headache. Headache arising de novo in a patient with known malignancy suggests either cerebral metastases or carcinomatous meningitis. Head pain appearing abruptly after bending, lifting, or coughing can be due to a posterior fossa mass, a Chiari malformation, or low cerebrospinal fluid (CSF) volume.

Brain tumors are discussed in Chap. 90.

■ TEMPORAL ARTERITIS (SEE ALSO CHAPS. 32 AND 363)

Temporal (giant cell) arteritis is an inflammatory disorder of arteries that frequently involves the extracranial carotid circulation. It is a common disorder of the elderly; its annual incidence is 77 per 100,000 individuals aged ≥50. The average age of onset is 70 years, and women account for 65% of cases. About half of patients with untreated temporal arteritis develop blindness due to involvement of the ophthalmic artery and its branches; indeed, the ischemic optic neuropathy induced by giant cell arteritis is the major cause of rapidly developing bilateral blindness in patients >60 years. Because treatment with glucocorticoids is effective in preventing this complication, prompt recognition of the disorder is important.

Typical presenting symptoms include headache, polymyalgia rheumatica (**Chap. 363**), jaw claudication, fever, and weight loss. Headache

TABLE 16-2 Headache Symptoms That Suggest a Serious Underlying Disorder

Sudden-onset headache
First severe headache
“Worst” headache ever
Vomiting that precedes headache
Subacute worsening over days or weeks
Pain induced by bending, lifting, coughing
Pain that disturbs sleep or presents immediately upon awakening
Known systemic illness
Onset after age 55
Fever or unexplained systemic signs
Abnormal neurologic examination
Pain associated with local tenderness, e.g., region of temporal artery

is the dominant symptom and often appears in association with malaise and muscle aches. Head pain may be unilateral or bilateral and is located temporally in 50% of patients but may involve any and all aspects of the cranium. Pain usually appears gradually over a few hours before peak intensity is reached; occasionally, it is explosive in onset. The quality of pain is infrequently throbbing; it is almost invariably described as dull and boring, with superimposed episodic stabbing pains similar to the sharp pains that appear in migraine. Most patients can recognize that the origin of their head pain is superficial, external to the skull, rather than originating deep within the cranium (the pain site usually identified by migraineurs). Scalp tenderness is present, often to a marked degree; brushing the hair or resting the head on a pillow may be impossible because of pain. Headache is usually worse at night and often aggravated by exposure to cold. Additional findings may include reddened, tender nodules or red streaking of the skin overlying the temporal arteries, and tenderness of the temporal or, less commonly, the occipital arteries.

The erythrocyte sedimentation rate (ESR) is often, although not always, elevated; a normal ESR does not exclude giant cell arteritis. A temporal artery biopsy followed by immediate treatment with prednisone 80 mg daily for the first 4–6 weeks should be initiated when clinical suspicion is high; treatment should not be unreasonably delayed to obtain a biopsy. The prevalence of migraine among the elderly is substantial, considerably higher than that of giant cell arteritis. Migraineurs often report amelioration of their headache with prednisone; thus, caution must be used when interpreting the therapeutic response.

■ GLAUCOMA

Glaucoma may present with a prostrating headache associated with nausea and vomiting. The headache often starts with severe eye pain. On physical examination, the eye is often red with a fixed, moderately dilated pupil.

Glaucoma is discussed in Chap. 32.

PRIMARY HEADACHE DISORDERS

Primary headaches are disorders in which headache and associated features occur in the absence of any exogenous cause. The most common are migraine, tension-type headache, and the TACs, notably cluster headache. These entities are discussed in detail in **Chap. 430**.

■ CHRONIC DAILY OR NEAR-DAILY HEADACHE

The broad description of chronic daily headache (CDH) can be applied when a patient experiences headache on 15 days or more per month. CDH is neither a single entity nor a diagnosis; it encompasses a number of different headache syndromes, both primary and secondary (**Table 16-3**). In aggregate, this group presents considerable

TABLE 16-3 Classification of Daily or Near-Daily Headache

Primary		
>4 H DAILY	<4 H DAILY	SECONDARY
Chronic migraine ^a	Chronic cluster headache ^b	Posttraumatic Head injury Iatrogenic Postinfectious
Chronic tension-type headache ^a	Chronic paroxysmal hemicrania	Inflammatory, such as Giant cell arteritis Sarcoidosis Behcet's syndrome
Hemicrania continua ^a	SUNCT/SUNA	Chronic CNS infection
New daily persistent headache ^a	Hypnic headache	Medication-overuse headache ^a

^aMay be complicated by medication overuse. ^bSome patients may have headache >4 h/d.

Abbreviations: CNS, central nervous system; SUNA, short-lasting unilateral neuralgiform headache attacks with cranial autonomic symptoms; SUNCT, short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing.

disability and is thus specially mentioned here. Population-based estimates suggest that about 4% of adults have daily or near-daily headache.

APPROACH TO THE PATIENT

Chronic Daily Headache

The first step in the management of patients with CDH is to diagnose any secondary headache and treat that problem (**Table 16-3**). This can sometimes be a challenge when the underlying cause triggers worsening of a primary headache. For patients with primary headaches, diagnosis of the headache type will guide therapy. Preventive treatments such as tricyclics, either amitriptyline or nortriptyline, at doses up to 1 mg/kg, are very useful in patients with CDH arising from migraine or tension-type headache or where the secondary cause has activated the underlying primary headache. Tricyclics are started in low doses (10–25 mg daily) and may be given 12 h before the expected time of awakening in order to avoid excessive morning sleepiness. Medicines including topiramate, valproate, propranolol, flunarizine (not available in the United States), candesartan, and the newer calcitonin gene-related peptide (CGRP) pathway monoclonal antibodies, or gepants-CGRP receptor antagonists (see **Chap. 430**) are also useful when the underlying issue is migraine.

MANAGEMENT OF MEDICALLY INTRACTABLE DISABLING PRIMARY HEADACHE

The management of medically intractable headache is difficult, although recent developments in therapy are at hand. Monoclonal antibodies to CGRP or its receptor have been reported to be effective and well tolerated in chronic migraine and are now licensed for use in clinical practice. Noninvasive neuromodulatory approaches, such as single-pulse transcranial magnetic stimulation and noninvasive vagal nerve stimulation, which appear to modulate thalamic processing or brainstem mechanisms, respectively, in migraine have been used in clinical practice with success. Noninvasive vagal nerve stimulation has also shown promise particularly in chronic cluster headache, chronic paroxysmal hemicrania, and hemicrania continua, and possibly in short-lasting unilateral neuralgiform headache attacks with cranial autonomic symptoms (SUNA) and short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing (SUNCT) (**Chap. 430**). Other modalities are discussed in **Chap. 430**.

MEDICATION-RELATED AND MEDICATION-OVERUSE HEADACHE

Overuse of analgesic medication for headache can aggravate headache frequency, markedly impair the effect of preventive medicines, and induce a state of refractory daily or near-daily headache called *medication-overuse headache*. A proportion of patients who stop taking analgesics will experience substantial improvement in the severity and frequency of their headache. However, even after cessation of analgesic use, many patients continue to have headache, although they may feel clinically improved in some way, especially if they have been using opioids or barbiturates regularly. The residual symptoms probably represent the underlying primary headache disorder, and most commonly this issue occurs in patients prone to migraine.

Management of Medication Overuse: Outpatients For patients who overuse analgesic medications, it is often helpful to reduce and eliminate the medications, although this approach is far from universally effective. One approach is to reduce the medication dose by 10% every 1–2 weeks. Immediate cessation of analgesic use is possible for some patients, provided there is no contraindication. Both approaches are facilitated by use of a medication diary maintained during the month or two before cessation; this helps to identify the scope of the problem. A small dose of a nonsteroidal anti-inflammatory drug (NSAID) such as naproxen, 500 mg bid, if tolerated, will help relieve residual pain as analgesic use is reduced.

NSAID overuse is not usually a problem for patients with daily headache when an NSAID with a longer half-life is taken once or twice daily; however, overuse problems may develop with shorter-acting NSAIDS. Once the patient has substantially reduced analgesic use, a preventive medication should be introduced. Another widely used approach is to commence the preventive at the same time the analgesic reduction is started. It must be emphasized that *preventives may not work in the presence of analgesic overuse, particularly with opioids*. The most common cause of unresponsiveness to treatment is the use of a preventive when analgesics continue to be used regularly. For some patients, discontinuing analgesics is very difficult; often the best approach is to inform the patient that some degree of headache is inevitable during this initial period.

Management of Medication Overuse: Inpatients Some patients will require hospitalization for detoxification. Such patients have typically failed efforts at outpatient withdrawal or have a significant medical condition, such as diabetes mellitus or epilepsy, which would complicate withdrawal as an outpatient. Following admission to the hospital, medications are withdrawn completely on the first day, in the absence of a contraindication. Antiemetics and fluids are administered as required; clonidine is used for opioid withdrawal symptoms. For acute intolerable pain during the waking hours, aspirin, 1 g IV (not approved in the United States), is useful. IM chlorpromazine can be helpful at night; patients must be adequately hydrated. Three to five days into the admission, as the effect of the withdrawn substance wears off, a course of IV dihydroergotamine (DHE) can be used. DHE, administered every 8 h for 5 consecutive days, a treatment that is not stopped short if headache settles, can induce a significant remission that allows a preventive treatment to be established. Serotonin 5-HT₃ receptor antagonists, such as ondansetron or granisetron, or the neurokinin receptor antagonist, aprepitant, may be required with DHE to prevent significant nausea, and domperidone (not approved in the United States) orally or by suppository can be very helpful. Avoiding sedating or otherwise side effect-prone antiemetics is helpful.

NEW DAILY PERSISTENT HEADACHE

New daily persistent headache (NDPH) is a clinically distinct syndrome with important secondary causes; these are listed in **Table 16-4**.

Clinical Presentation NDPH presents with headache on most if not all days, and the patient can clearly, and often vividly, recall the moment of onset. The headache usually begins abruptly, but onset may be more gradual; evolution over 3 days has been proposed as the upper limit for this syndrome. Patients typically recall the exact day and circumstances of the onset of headache; the new, persistent head pain does not remit. The first priority is to distinguish between a primary and a secondary cause of this syndrome. Subarachnoid hemorrhage is the most serious of the secondary causes and must be excluded either by history or appropriate investigation (**Chap. 429**).

Secondary NDPH • Low CSF Volume Headache In these syndromes, head pain is positional: it begins when the patient sits or stands upright and resolves upon reclining. The pain, which is occipitofrontal, is usually a dull ache but may be throbbing. Patients with chronic low CSF volume headache typically present with a

history of headache from one day to the next that is generally not present on waking but worsens during the day. Recumbency usually improves the headache within minutes, and it can take only minutes to an hour for the pain to return when the patient resumes an upright position.

The most common cause of headache due to persistent low CSF volume is CSF leak following LP (**Chap. S9**). Post-LP headache usually begins within 48 h but may be delayed for up to 12 days. Its incidence is between 10% and 30%. Beverages with caffeine may provide temporary relief. Besides LP, index events may include epidural injection or a vigorous Valsalva maneuver, such as from lifting, straining, coughing, clearing the eustachian tubes in an airplane, or multiple orgasms. Spontaneous CSF leaks are well recognized, and the diagnosis should be considered whenever the headache history is typical, even when there is no obvious index event. As time passes from the index event, the postural nature may become less apparent; cases in which the index event occurred several years before the eventual diagnosis have been recognized. Symptoms appear to result from low volume rather than low pressure: although low CSF pressures, typically 0–50 mm CSF, are usually identified, a pressure as high as 140 mm CSF has been noted with a documented leak.

Postural orthostatic tachycardia syndrome (POTS; **Chap. 440**) can present with orthostatic headache similar to low CSF volume headache and is a diagnosis that needs consideration in this setting.

When imaging is indicated to identify the source of a presumed leak, an MRI with gadolinium is the initial study of choice (**Fig. 16-1**). A striking pattern of diffuse meningeal enhancement is so typical that in the appropriate clinical context the diagnosis is established. Chiari malformations may sometimes be noted on MRI; in such cases, surgery to decompress the posterior fossa is *not* indicated and usually worsens the headache. Spinal MRI with T2 weighting may reveal a leak, and spinal MRI may demonstrate spinal meningeal cysts whose role in these syndromes is yet to be elucidated. The source of CSF leakage may be identified by spinal MRI with appropriate sequences, or by CT, preferably digital subtraction, myelography. In the absence of a directly identified site of leakage, ¹¹¹In-DTPA CSF studies may demonstrate early emptying of the tracer into the bladder or slow progress of tracer across the brain suggesting a CSF leak; this procedure is now only rarely employed.

TABLE 16-4 Differential Diagnosis of New Daily Persistent Headache

PRIMARY	SECONDARY
Migrainous-type	Subarachnoid hemorrhage
Featureless (tension-type)	Low cerebrospinal fluid (CSF) volume headache Raised CSF pressure headache Posttraumatic headache ^a Chronic meningitis

^aIncludes postinfectious forms.

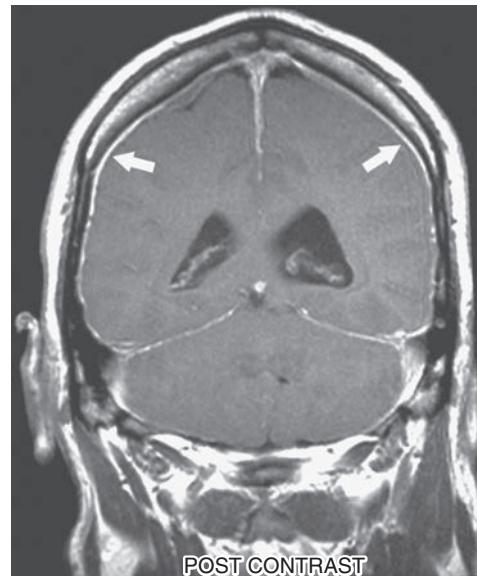


FIGURE 16-1 Magnetic resonance image showing diffuse meningeal enhancement after gadolinium administration in a patient with low cerebrospinal fluid (CSF) volume headache.

Initial treatment for low CSF volume headache is bed rest. For patients with persistent pain, IV caffeine (500 mg in 500 mL of saline administered over 2 h) can be very effective. An electrocardiogram (ECG) to screen for arrhythmia should be performed before administration. It is reasonable to administer at least two infusions of caffeine before embarking on additional tests to identify the source of the CSF leak. Because IV caffeine is safe and can be curative, it spares many patients the need for further investigations. If unsuccessful, an abdominal binder may be helpful. If a leak can be identified, an autologous blood patch is usually curative. A blood patch is also effective for post-LP headache; in this setting, the location is empirically determined to be the site of the LP. In patients with intractable headache, oral theophylline is a useful alternative that can take some months to be effective.

Raised CSF Pressure Headache Raised CSF pressure is well recognized as a cause of headache. Brain imaging can often reveal the cause, such as a space-occupying lesion.

Idiopathic intracranial hypertension (pseudotumor cerebri) NDPH due to raised CSF pressure can be the presenting symptom for patients with idiopathic intracranial hypertension, a disorder associated with obesity, female gender, and, on occasion, pregnancy. The syndrome can also occur without visual problems, particularly when the fundi are normal. These patients typically present with a history of generalized headache that is present on waking and improves as the day goes on. It is generally present on awakening in the morning and is worse with recumbency. Transient visual obscurations are frequent and may occur when the headaches are most severe. The diagnosis is relatively straightforward when papilledema is present, but the possibility must be considered even in patients without funduscopic changes. Formal visual field testing should be performed even in the absence of overt ophthalmic involvement. Partial obstructions of the cerebral venous sinuses are found in a small number of cases. In addition, persistently raised intracranial pressure can trigger a syndrome of chronic migraine. Other conditions that characteristically produce headache on rising in the morning or nocturnal headache are obstructive sleep apnea or poorly controlled hypertension.

Evaluation of patients suspected to have raised CSF pressure requires brain imaging. It is most efficient to obtain an MRI, including an MR venogram, as the initial study. If there are no contraindications, the CSF pressure should be measured by LP; this should be done when the patient is symptomatic so that both the pressure and the response to removal of 20–30 mL of CSF can be determined. An elevated opening pressure and improvement in headache following removal of CSF are diagnostic in the absence of fundal changes.

Initial treatment is with acetazolamide (250–500 mg bid); the headache may improve within weeks. If ineffective, topiramate is the next treatment of choice; it has many actions that may be useful in this setting, including carbonic anhydrase inhibition, weight loss, and neuronal membrane stabilization, likely mediated via effects on phosphorylation pathways. Severely disabled patients who do not respond to medical treatment require intracranial pressure monitoring and may require shunting. If appropriate, weight loss should be encouraged.

Posttraumatic Headache A traumatic event can trigger a headache process that lasts for many months or years after the event. The term *trauma* is used here in a very broad sense: headache can develop following an injury to the head, but it can also develop after an infectious episode, typically viral meningitis; a flulike illness; or a parasitic infection. Complaints of dizziness, vertigo, and impaired memory can accompany the headache. Symptoms may remit after several weeks or persist for months and even years after the injury. Typically, the neurologic examination is normal and CT or MRI studies are unrevealing. Chronic subdural hematoma may

on occasion mimic this disorder. Posttraumatic headache may also be seen after carotid dissection and subarachnoid hemorrhage and after intracranial surgery. The underlying theme appears to be that a traumatic event involving the pain-producing meninges can trigger a headache process that lasts for many years.

Other Causes In one series, one-third of patients with NDPH reported headache beginning after a transient flulike illness characterized by fever, neck stiffness, photophobia, and marked malaise. Evaluation typically reveals no apparent cause for the headache. There is no convincing evidence that persistent Epstein-Barr virus infection plays a role in NDPH. A complicating factor is that many patients undergo LP during the acute illness; iatrogenic low CSF volume headache must be considered in these cases.

Treatment Treatment is largely empirical and directed at the headache phenotype. Tricyclic antidepressants, notably amitriptyline, and anticonvulsants, such as topiramate, valproate, candesartan, and gabapentin, have been used with reported benefit. The monoamine oxidase inhibitor phenelzine may also be useful in carefully selected patients. The headache usually resolves within 3–5 years, but it can be quite disabling.

PRIMARY CARE AND HEADACHE MANAGEMENT

Most patients with headache will be seen first in a primary care setting. The challenging task of the primary care physician is to identify the very few worrisome secondary headaches from the very great majority of primary and less dangerous secondary headaches (Table 16-2).

Absent any warning signs, a reasonable approach is to treat when a diagnosis is established. As a general rule, the investigation should focus on identifying worrisome causes of headache or on helping the patient to gain confidence if no primary headache diagnosis can be made.

After treatment has been initiated, follow-up care is essential to identify whether progress has been made against the headache complaint. Not all headaches will respond to treatment, but, in general, worrisome headaches will progress and will be easier to identify.

When a primary care physician feels the diagnosis is a primary headache disorder, it is worth noting that >90% of patients who present to primary care with a complaint of headache will have migraine (Chap. 430).

In general, patients who do not have a clear diagnosis, have a primary headache disorder other than migraine or tension-type headache, or are unresponsive to two or more standard therapies for the considered headache type, should be considered for referral to a specialist. In a practical sense, the threshold for referral is also determined by the experience of the primary care physician in headache medicine and the availability of secondary care options.

ACKNOWLEDGMENT

The editors acknowledge the contributions of Neil H. Raskin to earlier editions of this chapter.

FURTHER READING

- HEADACHE CLASSIFICATION COMMITTEE OF THE INTERNATIONAL HEADACHE SOCIETY: The International Classification of Headache Disorders, 3rd ed. *Cephalgia* 33:629, 2018.
- KERNICK D, GOADSBY PJ: *Headache: A Practical Manual*. Oxford: Oxford University Press, 2008.
- LANCE JW, GOADSBY PJ: *Mechanism and Management of Headache*, 7th ed. New York, Elsevier, 2005.
- OLESEN J et al: *The Headaches*. Philadelphia, Lippincott, Williams & Wilkins, 2005.
- SILBERSTEIN SD, LIPTON RB, DODICK DW: *Wolff's Headache and Other Head Pain*, 9th ed. New York, Oxford University Press, 2021.



The importance of back and neck pain in our society is underscored by the following: (1) the cost of chronic back pain in the United States is estimated at more than \$200 billion annually; approximately one-third of this cost is due to direct health care expenses and two-thirds are indirect costs resulting from loss of wages and productivity; (2) back symptoms are the most common cause of disability in individuals <45 years of age; (3) low back pain (LBP) is the second most common reason for visiting a physician in the United States; and (4) more than four out of five people will experience significant back pain at some point in their lives.

ANATOMY OF THE SPINE

The anterior spine consists of cylindrical vertebral bodies separated by intervertebral disks and stabilized by the anterior and posterior longitudinal ligaments. The intervertebral disks are composed of a central gelatinous nucleus pulposus surrounded by a tough cartilaginous ring, the annulus fibrosis. Disks are responsible for 25% of spinal column length and allow the bony vertebrae to move easily upon each other (Figs. 17-1 and 17-2). Desiccation of the nucleus pulposus and degeneration of the annulus fibrosus worsen with age, resulting in loss of disk height. The disks are largest in the cervical and lumbar regions where movements of the spine are greatest. The anterior spine absorbs the shock of bodily movements such as walking and running, and with the posterior spine protects the spinal cord and nerve roots in the spinal canal.

The posterior spine consists of the vertebral arches and processes. Each arch consists of paired cylindrical pedicles anteriorly and paired lamina posteriorly. The vertebral arch also gives rise to two transverse processes laterally, one spinous process posteriorly, plus two superior and two inferior articular facets. The apposition of a superior and inferior facet constitutes a *facet joint*. The posterior spine provides an anchor for the attachment of muscles and ligaments. The contraction of muscles attached to the spinous and transverse processes and lamina works like a system of pulleys and levers producing flexion, extension, rotation, and lateral bending movements of the spine.

Nerve root injury (*radiculopathy*) is a common cause of pain in the neck and arm, or low back and buttock, or leg (see **dermatomes** in Figs. 25-2 and 25-3). Each nerve root exits just above its corresponding vertebral body in the cervical region (e.g., the C7 nerve root exits

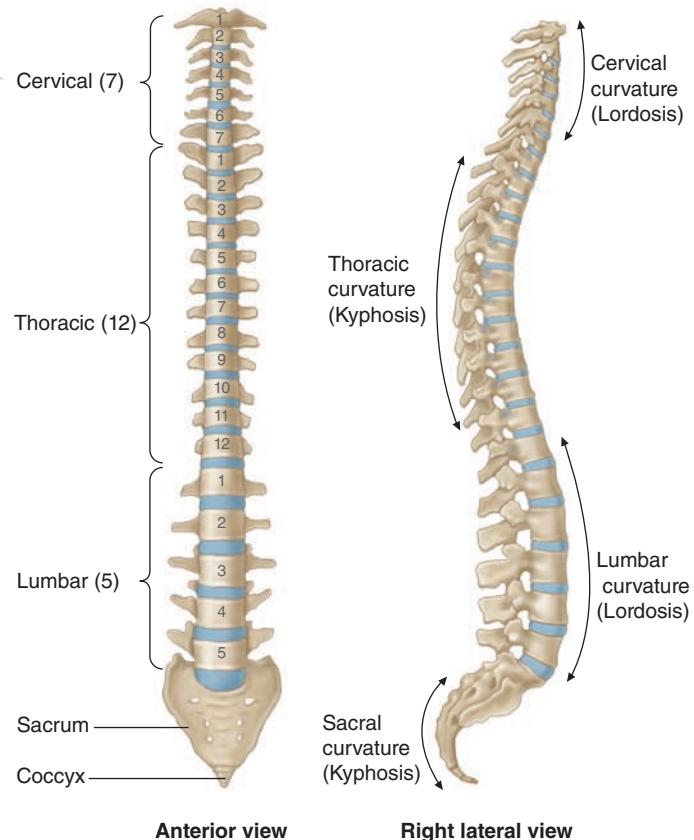


FIGURE 17-2 Spinal column. (Reproduced with permission from AG Cornuelle, DH Gronefeld: Radiographic Anatomy Positioning. New York, McGraw-Hill, 1998.)

at the C6-C7 level), and just below the vertebral body in the thoracic and lumbar spine (e.g., the T1 nerve root exits at the T1-T2 level). The cervical nerve roots follow a short intraspinal course before exiting. In contrast, because the spinal cord ends at the L1 or L2 vertebral level, the lumbar nerve roots follow a long intraspinal course and can be injured anywhere along its path. For example, disk herniation at the L4-L5 level can produce L4 root compression laterally, but more often compression of the traversing L5 nerve root occurs (Fig. 17-3). The lumbar nerve roots are mobile in the spinal canal, but eventually pass through the narrow *lateral recess* of the spinal canal and *intervertebral*

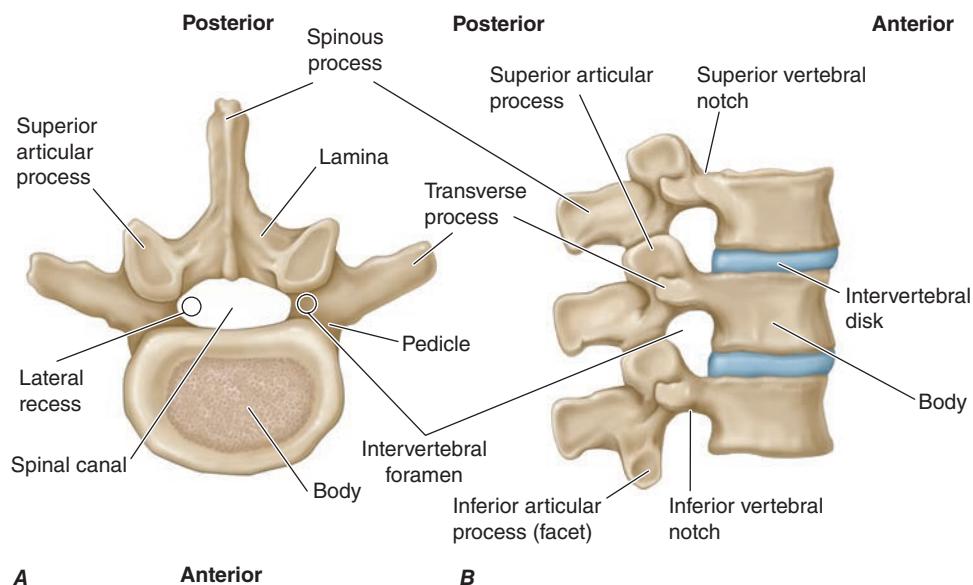


FIGURE 17-1 Vertebral anatomy. **A.** Vertebral body—axial view; **B.** vertebral column—sagittal view. (Reproduced with permission from AG Cornuelle, DH Gronefeld: Radiographic Anatomy Positioning. New York, McGraw-Hill, 1998.)

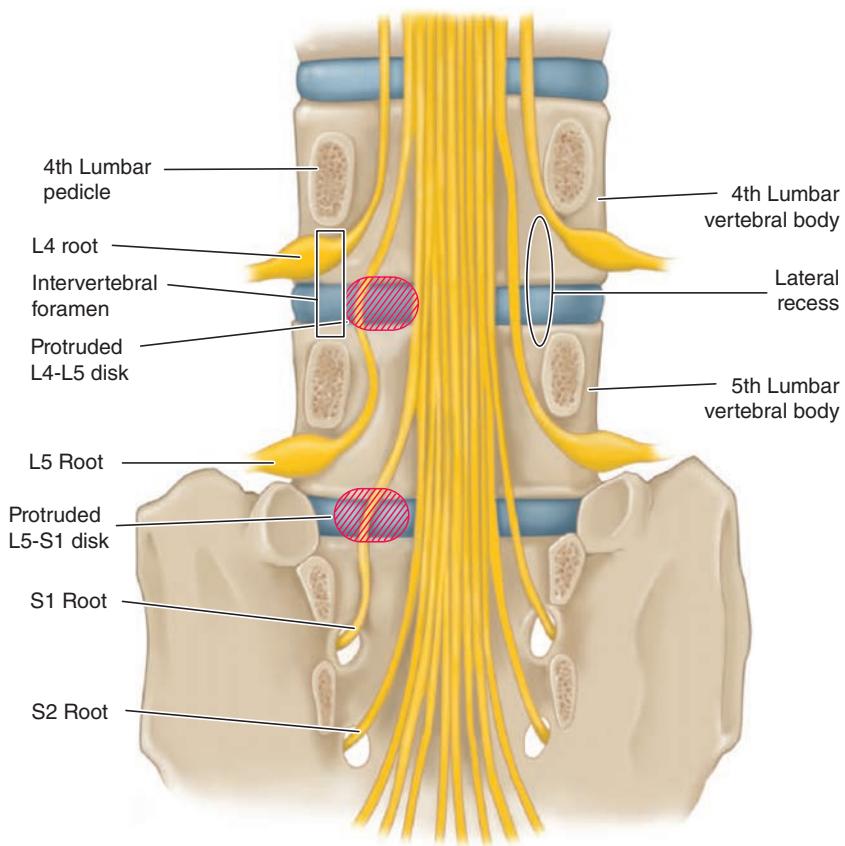


FIGURE 17-3 Compression of L5 and S1 roots by herniated disks. (Reproduced with permission from AH Ropper, MA Samuels: Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009.)

foramen (Figs. 17-2 and 17-3). When imaging the spine, both sagittal and axial views are needed to assess possible compression at these sites.

Beginning at the C3 level, each cervical (and the first thoracic) vertebral body projects a lateral bony process upward—the uncinate process. The uncinate process articulates with the cervical vertebral body above via the uncovertebral joint. The uncovertebral joint can hypertrophy with age and contribute to neural foraminal narrowing and cervical radiculopathy.

Pain-sensitive structures of the spine include the periosteum of the vertebrae, dura, facet joints, annulus fibrosus of the intervertebral disk, epidural veins and arteries, and the longitudinal ligaments. Disease of these diverse structures may explain many cases of back pain without nerve root compression. Under normal circumstances, the nucleus pulposus of the intervertebral disk is not pain sensitive.

Pain of spine origin may be located in the back or referred to the buttocks or legs. Diseases affecting the upper lumbar spine tend to refer pain to the lumbar region, groin, or anterior thighs. Diseases affecting the lower lumbar spine tend to produce pain referred to the buttocks, posterior thighs, calves, or feet. Referred pain often explains pain syndromes that cross multiple dermatomes without evidence of nerve or nerve root injury.

APPROACH TO THE PATIENT

Back Pain

TYPES OF BACK PAIN

Delineating the type of pain reported by the patient is the essential first step. Attention is also focused on identifying risk factors for a serious underlying etiology. The most frequent serious causes of back pain are radiculopathy, fracture, tumor, infection, or referred pain from visceral structures (Table 17-1).

Local pain is caused by injury to pain-sensitive structures that compress or irritate sensory nerve endings. The site of the pain is near the affected part of the back.

Pain referred to the back may arise from abdominal or pelvic viscera. The pain is usually described as primarily abdominal or pelvic, accompanied by back pain, and usually unaffected by posture. The patient may occasionally complain of back pain only.

TABLE 17-1 Acute Low Back Pain: Risk Factors for an Important Structural Cause

History

- Pain worse at rest or at night
- Prior history of cancer
- History of chronic infection (especially lung, urinary tract, skin, poor dentition)
- History of trauma
- Incontinence
- Age >70 years
- Intravenous drug use
- Glucocorticoid use
- History of a rapidly progressive neurologic deficit

Examination

- Unexplained fever
- Unexplained weight loss
- Focal palpation/percussion tenderness over the midline spine
- Abdominal, rectal, or pelvic mass
- Internal/external rotation of the leg at the hip
- Straight-leg or reverse straight-leg raising signs
- Progressive focal neurologic deficit

Radicular pain is typically sharp and radiates from the low back to a leg within the territory of a nerve root (see “Lumbar Disk Disease,” below). Coughing, sneezing, or voluntary contraction of abdominal muscles (lifting heavy objects or straining at stool) may elicit or worsen the radiating pain. The pain may also increase in postures that stretch the nerves and nerve roots. Sitting with the leg outstretched places traction on the sciatic nerve and L5 and S1 roots because the sciatic nerve passes posterior to the hip. The femoral nerve (L2, L3, and L4 roots) passes anterior to the hip and is not stretched by sitting. The description of the pain alone often fails to distinguish between referred pain and radiculopathy, although a burning or electric quality favors radiculopathy.

Pain associated with muscle spasm is commonly associated with many spine disorders. The spasms may be accompanied by an abnormal posture, tense paraspinal muscles, and dull or achy pain in the paraspinal region.

Knowledge of the circumstances associated with the onset of back pain is important when weighing possible serious underlying causes for the pain. Some patients involved in accidents or work-related injuries may exaggerate their pain for the purpose of compensation or for psychological reasons.

EXAMINATION

A complete physical examination including vital signs, heart and lungs, abdomen and rectum, and limbs is advisable. Back pain referred from visceral organs may be reproduced during palpation of the abdomen (pancreatitis, abdominal aortic aneurysm [AAA]) or percussion over the costovertebral angles (pyelonephritis).

The normal spine has a cervical and lumbar lordosis and a thoracic kyphosis. Exaggeration of these normal alignments may result in hyperkyphosis of the thoracic spine or hyperlordosis of the lumbar spine. Inspection of the back may reveal a lateral curvature of the spine (scoliosis). A midline hair tuft, skin dimpling or pigmentation, or a sinus tract may indicate a congenital spine anomaly. Asymmetry in the prominence of the paraspinal muscles suggests muscle spasm. Palpation over the spinous process transmits force to the entire vertebrae and suggests vertebral pathology.

Flexion at the hips is normal in patients with lumbar spine disease, but flexion of the lumbar spine is limited and sometimes painful. Lateral bending to the side opposite the injured spinal element may stretch the damaged tissues, worsen pain, and limit motion. Hyperextension of the spine (with the patient prone or standing) is limited when nerve root compression, facet joint pathology, or other bony spine disease is present.

Pain from hip disease may mimic the pain of lumbar spine disease. Hip pain can be reproduced by passive internal and external rotation at the hip with the knee and hip in flexion or by percussing the heel with the examiner’s palm with the leg extended (heel percussion sign).

The *straight-leg raising (SLR)* maneuver is a simple bedside test for nerve root disease. With the patient supine, passive straight-leg flexion at the hip stretches the L5 and S1 nerve roots and the sciatic nerve; dorsiflexion of the foot during the maneuver adds to the stretch. In healthy individuals, flexion to at least 80° is normally possible without causing pain, although a tight, stretching sensation in the hamstring muscles is common. The SLR test is positive if the maneuver reproduces the patient’s usual back or limb pain. Eliciting the SLR sign in both the supine and sitting positions can help determine if the finding is reproducible. The patient may describe pain in the low back, buttocks, posterior thigh, or lower leg, but the *key feature is reproduction of the patient’s usual pain*. The *crossed SLR sign* is present when flexion of one leg reproduces the usual pain in the opposite leg or buttocks. In disk herniation, the crossed SLR sign is less sensitive but more specific than the SLR sign. The *reverse SLR sign* is elicited by standing the patient next to the examination table and passively extending each leg with the knee fully extended. This maneuver, which stretches the L2-L4 nerve roots, lumbosacral plexus, and femoral nerve, is considered positive if the patient’s usual back or limb pain is reproduced. For all of these tests, the nerve or nerve root lesion is always on the side of the pain. Examination of the unaffected leg first provides a control test, ensures mutual understanding of test parameters, and enhances test utility.

The neurologic examination includes a search for focal weakness or muscle atrophy, localized reflex changes, diminished sensation in the legs, or signs of spinal cord injury. The examiner should be alert to the possibility of breakaway weakness, defined as fluctuations in the maximum power generated during muscle testing. Breakaway weakness may be due to pain, inattention, or a combination of pain and underlying true weakness. Breakaway weakness without pain is usually due to a lack of effort. In uncertain cases, electromyography (EMG) can determine if true weakness due to nerve tissue injury is present. Findings with specific lumbosacral nerve root lesions are shown in **Table 17-2** and are discussed below.

LABORATORY, IMAGING, AND EMG STUDIES

Laboratory studies are rarely needed for the initial evaluation of nonspecific acute (<3 months duration) low back pain (ALBP).

TABLE 17-2 Lumbosacral Radiculopathy: Neurologic Features

LUMBOSACRAL NERVE ROOT	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
L2 ^a	—	Upper anterior thigh	Psoas (hip flexors)	Anterior thigh
L3 ^a	—	Lower anterior thigh Anterior knee	Psoas (hip flexors) Quadriceps (knee extensors) Thigh adductors	Anterior thigh, knee
L4 ^a	Quadriceps (knee)	Medial calf	Quadriceps (knee extensors) ^b Thigh adductors	Knee, medial calf Anterolateral thigh
L5 ^c	—	Dorsal surface—foot Lateral calf	Peronei (foot evertors) ^b Tibialis anterior (foot dorsiflexors) Gluteus medius (leg abductors) Toe dorsiflexors	Lateral calf, dorsal foot, posterior lateral thigh, buttocks
S1 ^c	Gastrocnemius/ soleus (ankle)	Plantar surface—foot Lateral aspect—foot	Gastrocnemius/soleus (foot plantar flexors) ^b Abductor hallucis (toe flexors) ^b Gluteus maximus (leg extensors)	Bottom foot, posterior calf, posterior thigh, buttocks

^aReverse straight-leg raising sign may be present—see “Examination of the Back.” ^bThese muscles receive the majority of innervation from this root. ^cStraight-leg raising sign may be present—see “Examination of the Back.”

Risk factors for a serious underlying cause and for infection, tumor, or fracture in particular should be sought by history and examination. If risk factors are present (Table 17-1), then laboratory studies (complete blood count [CBC], erythrocyte sedimentation rate [ESR], urinalysis) are indicated. If risk factors are absent, then management is conservative (see "Treatment," below).

CT scanning is used as a primary screening modality for acute trauma that is moderate to severe. CT is superior to x-rays for detection of fractures involving posterior spine structures, craniocervical and cervicothoracic junctions, C1 and C2 vertebrae, bone fragments in the spinal canal, or misalignment. MRI or CT myelography is the radiologic test of choice for evaluation of most serious diseases involving the spine. MRI is superior for the definition of soft tissue structures, whereas CT myelography provides optimal imaging of the lateral recess of the spinal canal, defines bony abnormalities, and is tolerated by claustrophobic patients.

Population surveys in the United States suggest that patients with back pain report greater functional limitations in recent years, despite rapid increases in spine imaging, opioid prescribing, injections, and spine surgery. This suggests that more selective use of diagnostic and treatment modalities may be reasonable for many patients. One prospective case-control study found that older adults with back pain of less than 6 weeks duration who received spine imaging as part of a primary care visit had no better outcomes than the control group.

Spine imaging often reveals abnormalities of dubious clinical relevance that may alarm clinicians and patients alike and prompt further testing and unnecessary therapy. When imaging tests are reviewed, it is important to remember that degenerative findings are common in normal, pain-free individuals. Randomized trials and observational studies have suggested that imaging can have a "cascade effect," creating a gateway to other unnecessary care. Interventions have included physician education and computerized decision support within the electronic medical record to require specific indications for approval of imaging tests. Other strategies have included audit and feedback of individual practitioners' rates of ordering, more rapid access to physical therapy, or consultation with spine experts for patients without imaging indications.

Educational tools created by the American College of Physicians for patients and the public have included "Five Things Physicians and Patients Should Question": (1) Do not recommend advanced imaging (e.g., MRI) of the spine within the first 6 weeks in patients with nonspecific ALBP in the absence of red flags. (2) Do not perform elective spinal injections without imaging guidance, unless contraindicated. (3) Do not use bone morphogenetic protein (BMP) for routine anterior cervical spine fusion surgery. (4) Do not use EMG and nerve conduction studies (NCSs) to determine the cause of purely midline lumbar, thoracic, or cervical spine pain. (5) Do not recommend bed rest for >48 h when treating LBP. In an observational study, application of this strategy was associated with lower rates of repeat imaging, opioid use, and referrals for physical therapy.

Electrodiagnostic studies can be used to assess the functional integrity of the peripheral nervous system (Chap. 446). Sensory NCSs are normal when focal sensory loss confirmed by examination is due to nerve root damage because the nerve roots are proximal to the nerve cell bodies in the dorsal root ganglia. Injury to nerve tissue distal to the dorsal root ganglion (e.g., plexus or peripheral nerve) results in reduced sensory nerve signals. Needle EMG complements NCSs by detecting denervation or reinnervation changes in a myotomal (segmental) distribution. Multiple muscles supplied by different nerve roots and nerves are sampled; the pattern of muscle involvement indicates the nerve root(s) responsible for the injury. Needle EMG provides objective information about motor nerve fiber injury when clinical evaluation of weakness is limited by pain or poor effort. EMG and NCSs will be normal when sensory nerve root injury or irritation is the pain source.

The COVID-19 pandemic has disrupted and complicated the care of patients with LBP. Paraspinal myalgias may result in LBP. The sedentary lifestyle resulting from quarantine is associated with an increased frequency or severity of LBP. Fear of infection risk has also prevented many patients from seeking needed care. Video-telemedicine visits can help identify patients with underlying risks for a serious cause and inform appropriate next steps in management.

CAUSES OF BACK PAIN (TABLE 17-3)

LUMBAR DISK DISEASE

Lumbar disk disease is a common cause of acute, chronic, or recurrent low back and leg pain (Figs. 17-3 and 17-4). Disk disease is most likely to occur at the L4-L5 or L5-S1 levels, but upper lumbar levels can also be involved. The cause is often unknown, but the risk is increased in overweight individuals. Disk herniation is unusual prior to age 20 years and is rare in the fibrotic disks of the elderly. Complex genetic factors may play a role in predisposition. The pain may be located in the low back only or referred to a leg, buttock, or hip. A sneeze, cough, or trivial movement may cause the nucleus pulposus to prolapse, pushing the frayed and weakened annulus posteriorly. With severe disk disease, the nucleus can protrude through the annulus (herniation) or become extruded to lie as a free fragment in the spinal canal.

TABLE 17-3 Causes of Back or Neck Pain

Lumbar or Cervical Disk Disease

Degenerative Spine Disease

Lumbar spinal stenosis without or with neurogenic claudication

Intervertebral foraminal or lateral recess narrowing

Disk-osteophyte complex

Facet or uncovertebral joint hypertrophy

Lateral disk protrusion

Spondylosis (osteoarthritis), spondylolisthesis, or spondylolysis

Spine Infection

Vertebral osteomyelitis

Spinal epidural abscess

Septic disk (diskitis)

Meningitis

Lumbar arachnoiditis

Neoplasms

Metastatic with/without pathologic fracture

Primary Nervous System: Meningioma, neurofibroma, schwannoma

Primary Bone: chordoma, osteoma

Trauma

Strain or sprain

Whiplash injury

Trauma/falls, motor vehicle accidents

Metabolic Spine Disease

Osteoporosis with/without pathologic fracture—hyperparathyroidism, immobility

Osteosclerosis (e.g., Paget's disease)

Congenital/Developmental

Spondylolysis

Kyphoscoliosis

Spina bifida occulta

Tethered spinal cord

Autoimmune Inflammatory Arthritis

Other Causes of Back Pain

Referred pain from visceral disease (e.g., abdominal aortic aneurysm)

Postural

Psychiatric, malingering, chronic pain syndromes

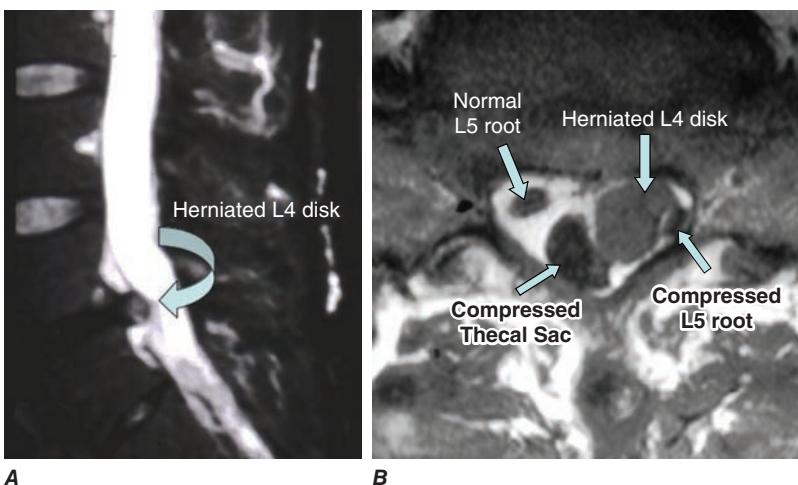


FIGURE 17-4 Disk herniation. **A.** Sagittal T2-weighted image on the left side of the spinal canal reveals disk herniation at the L4-L5 level. **B.** Axial T1-weighted image shows paracentral disk herniation with displacement of the thecal sac medially and the left L5 nerve root posteriorly in the left lateral recess.

The mechanism by which intervertebral disk injury causes back pain is uncertain. The inner annulus fibrosus and nucleus pulposus are normally devoid of innervation. Inflammation and production of proinflammatory cytokines within a ruptured nucleus pulposus may trigger or perpetuate back pain. Ingrowth of nociceptive (pain) nerve fibers into the nucleus pulposus of a diseased disk may be responsible for some cases of chronic “diskogenic” pain. Nerve root injury (radiculopathy) from disk herniation is usually due to inflammation, but lateral herniation may produce compression in the lateral recess or intervertebral foramen.

A ruptured disk may be asymptomatic or cause back pain, limited spine motion (particularly flexion), a focal neurologic deficit, or radicular pain. A dermatomal pattern of sensory loss or a reduced or absent deep tendon reflex is more suggestive of a specific root lesion than is the pattern of pain. Motor findings (focal weakness, muscle atrophy, or fasciculations) occur less frequently than focal sensory or reflex changes. Symptoms and signs are usually unilateral, but bilateral involvement does occur with large central disk herniations that involve roots bilaterally or cause inflammation of nerve roots within the spinal canal. Clinical manifestations of specific nerve root lesions are summarized in Table 17-2.

The differential diagnosis covers a variety of serious and treatable conditions, including epidural abscess, hematoma, fracture, or tumor. Fever, constant pain uninfluenced by position, sphincter abnormalities, or signs of myelopathy suggest an etiology other than lumbar disk disease. Absent ankle reflexes can be a normal finding in persons >60 years or a sign of bilateral S1 radiculopathies. An absent deep tendon reflex or focal sensory loss may indicate injury to a nerve root, but other sites of injury along the nerve must also be considered. As examples, an absent knee reflex may be due to a femoral neuropathy or an L4 nerve root injury; loss of sensation over the foot and lateral lower calf may result from a peroneal or lateral sciatic neuropathy, or an L5 nerve root injury. Focal muscle atrophy may reflect injury to the anterior horn cells of the spinal cord, a nerve root, peripheral nerve, or disuse.

A lumbar spine MRI scan or CT myelogram can often confirm the location and type of pathology. Spine MRIs yield exquisite views of intraspinal and adjacent soft tissue anatomy, whereas bony lesions of the lateral recess or intervertebral foramen are optimally visualized by CT myelography. The correlation of neuroradiologic findings to clinical symptoms, particularly pain, is not simple. Contrast-enhancing tears in the annulus fibrosus or disk protrusions are widely accepted as common sources of back pain; however, studies have found that many asymptomatic adults have similar radiologic findings. Entirely asymptomatic disk protrusions are also common, occurring in up to one-third of adults, and these may also enhance with contrast. Furthermore,

in patients with known disk herniation treated either medically or surgically, persistence of the herniation 10 years later had no relationship to the clinical outcome. In summary, MRI findings of disk protrusion, tears in the annulus fibrosus, or hypertrophic facet joints are common incidental findings that, by themselves, should not dictate management decisions for patients with back pain.

The diagnosis of nerve root injury is most secure when the history, examination, results of imaging studies, and the EMG are concordant. There is often good correlation between CT and EMG findings for localization of nerve root injury.

Management of lumbar disk disease is discussed below.

Cauda equina syndrome (CES) signifies an injury of multiple lumbosacral nerve roots within the spinal canal distal to the termination of the spinal cord at L1-L2. LBP, weakness and areflexia in the legs, saddle anesthesia, or loss of bladder function may occur. The problem must be distinguished from disorders of the lower spinal cord (conus medullaris syndrome), acute transverse myelitis (Chap. 442), and Guillain-Barré syndrome (Chap. 447).

Combined involvement of the conus medullaris and cauda equina can occur. CES is most commonly due to a large ruptured lumbosacral intervertebral disk, but other causes include lumbosacral spine fracture, hematoma within the spinal canal (sometimes following lumbar puncture in patients with coagulopathy), and tumor or other compressive mass lesions. Treatment is usually surgical decompression, sometimes on an urgent basis in an attempt to restore or preserve motor or sphincter function, or radiotherapy for metastatic tumors (Chap. 90).

■ DEGENERATIVE CONDITIONS

Lumbar spinal stenosis (LSS) describes a narrowed lumbar spinal canal. *Neurogenic claudication* consists of pain, typically in the back and buttocks or legs, that is brought on by walking or standing and relieved by sitting. Unlike vascular claudication, symptoms are often provoked by standing without walking. Unlike lumbar disk disease, symptoms are usually relieved by sitting. Patients with neurogenic claudication can often walk much farther when leaning over a shopping cart and can pedal a stationary bike with ease while sitting. These flexed positions increase the anteroposterior spinal canal diameter and reduce intraspinal venous hypertension, producing pain relief. Focal weakness, sensory loss, or reflex changes may occur when spinal stenosis is associated with neural foraminal narrowing and radiculopathy. Severe neurologic deficits, including paralysis and urinary incontinence, occur only rarely.

LSS by itself is common (6–7% of adults) and is usually asymptomatic. Symptoms are correlated with severe spinal canal stenosis. LSS is most often acquired (75%) but can also be congenital or due to a mixture of both etiologies. Congenital forms (achondroplasia and idiopathic) are characterized by short, thick pedicles that produce both spinal canal and lateral recess stenosis. Acquired factors that contribute to spinal stenosis include degenerative diseases (spondylosis, spondylolisthesis, and scoliosis), trauma, spine surgery, metabolic or endocrine disorders (epidural lipomatosis, osteoporosis, acromegaly, renal osteodystrophy, and hypoparathyroidism), and Paget's disease. MRI provides the best definition of the abnormal anatomy (Fig. 17-5).

LSS accompanied by neurogenic claudication responds to surgical decompression of the stenotic segments. The same processes leading to LSS may cause lumbar foraminal or lateral recess narrowing resulting in coincident lumbar radiculopathy that may require treatment as well.

Conservative treatment of symptomatic LSS can include nonsteroidal anti-inflammatory drugs (NSAIDs), acetaminophen, exercise programs, and symptomatic treatment of acute pain episodes. There is insufficient evidence to support the routine use of epidural glucocorticoid injections. Surgery is considered when medical therapy does not relieve symptoms sufficiently to allow for resumption of activities of

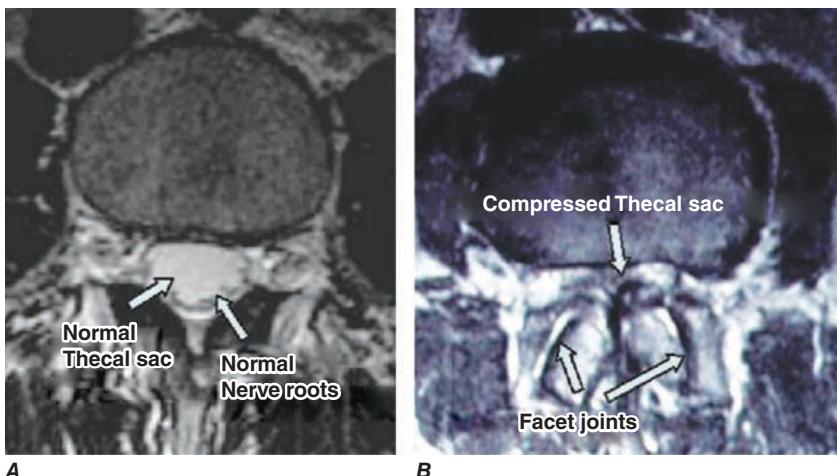


FIGURE 17-5 Spinal stenosis. **A**, An axial T2-weighted image of the normal lumbar spine shows a normal thecal sac within the lumbar spinal canal. The thecal sac is bright. The lumbar roots are seen as dark punctate dots located posteriorly in the thecal sac. **B**, The thecal sac is not well visualized due to severe lumbar spinal canal stenosis, partially the result of hypertrophic facet joints.

daily living or when focal neurologic signs are present. Most patients with neurogenic claudication who are treated medically do not improve over time. Surgical management with laminectomy, which increases the spinal canal diameter and reduces venous hypertension, can produce significant relief of exertional back and leg pain, leading to less disability and improved functional outcomes. Laminectomy and fusion is usually reserved for patients with LSS and spondylolisthesis. Predictors of a poor surgical outcome include impaired walking pre-operatively, depression, cardiovascular disease, and scoliosis. Up to one-quarter of surgically treated patients develop recurrent stenosis at the same or an adjacent spinal level within 7–10 years; recurrent symptoms usually respond to a second surgical decompression.

Neural foraminal narrowing or lateral recess stenosis with radiculopathy is a common consequence of osteoarthritic processes that cause LSS (Figs. 17-1 and 17-6), including osteophytes, lateral disk protrusion, calcified disk-osteophytes, facet joint hypertrophy, uncovertebral

joint hypertrophy (in the cervical spine), congenitally shortened pedicles, or, frequently, a combination of these processes. Neoplasms (primary or metastatic), fractures, infections (epidural abscess), or hematomas are less frequent causes. Most common is bony foraminal narrowing leading to nerve root ischemia and persistent symptoms, in contrast to inflammation that is associated with a paracentral herniated disk and radiculopathy. These conditions can produce unilateral nerve root symptoms or signs due to compression at the intervertebral foramen or in the lateral recess; symptoms are indistinguishable from disk-related radiculopathy, but treatment may differ depending on the etiology. The history and neurologic examination alone cannot distinguish between these possibilities. Neuroimaging (CT or MRI) is required to identify the anatomic cause. Neurologic findings from the examination and EMG can help direct the attention of the radiologist to specific nerve roots, especially on axial images. For *facet joint hypertrophy with foraminal stenosis*, surgical foraminotomy produces long-term relief of leg and back pain in 80–90% of patients. Facet joint or medial branch blocks for back or neck

pain are sometimes used to help determine the anatomic origin of back pain or for treatment, but there is a lack of clinical data to support their utility. Medical causes of lumbar or cervical radiculopathy unrelated to primary spine disease include infections (e.g., herpes zoster and Lyme disease), carcinomatous meningitis, diabetes, and root avulsion or traction (trauma).

Spondylosis and Spondylolisthesis

Spondylosis, or osteoarthritic spine disease, typically occurs in later life and primarily involves the cervical and lumbosacral spine. Patients often complain of back pain that increases with movement, is associated with stiffness, and is better with inactivity. The relationship between clinical symptoms and radiologic findings is usually not straightforward. Pain may be prominent when MRI, CT, or x-ray findings are minimal, and prominent degenerative spine disease can be seen in asymptomatic patients. Osteophytes, combined

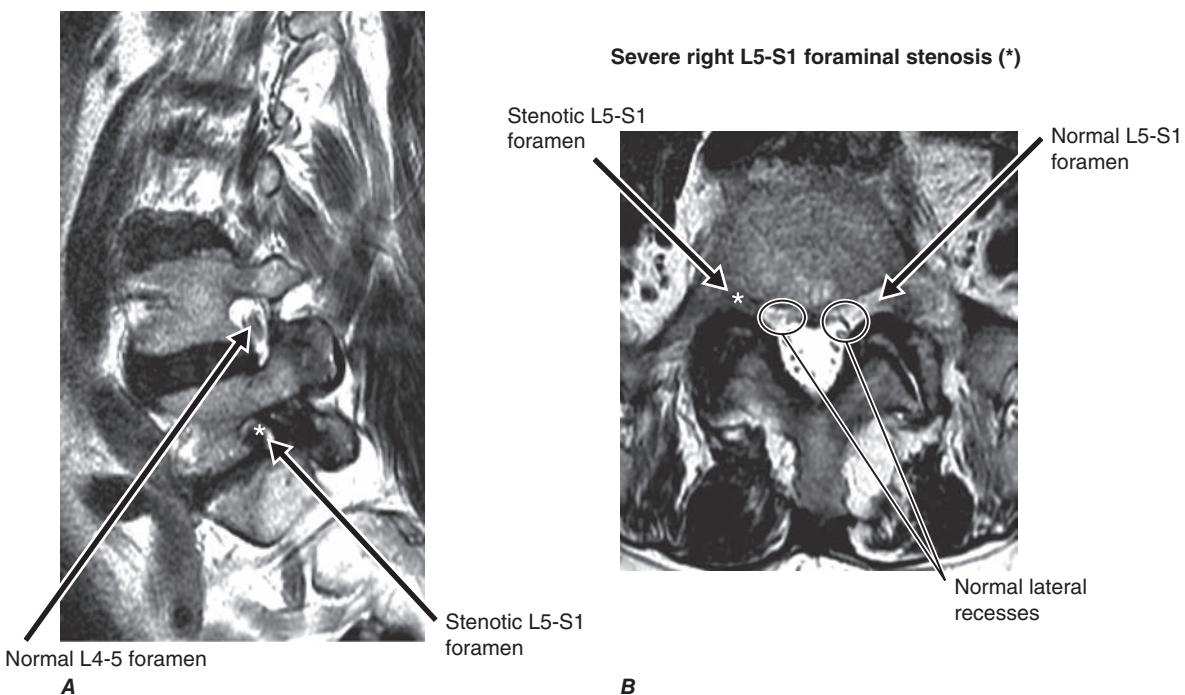


FIGURE 17-6 Foraminal stenosis. **A**, Sagittal T2-weighted image reveals normal high signal around the exiting right L4 nerve root in the right neural foramen at L4-L5; effacement of the high signal is noted one level below at L5-S1, due to severe foraminal stenosis. **B**, Axial T2-weighted image at the L5-S1 level demonstrates normal lateral recesses bilaterally, a normal intervertebral foramen on the left, but a severely stenotic foramen (*) on the right.

disk-osteophytes, or a thickened ligamentum flavum may cause or contribute to central spinal canal stenosis, lateral recess stenosis, or neural foraminal narrowing.

Spondylolisthesis is the anterior slippage of the vertebral body, pedicles, and superior articular facets, leaving the posterior elements behind. Spondylolisthesis can be associated with spondylolysis, congenital anomalies, degenerative spine disease, or other causes of mechanical weakness of the pars interarticularis (e.g., infection, osteoporosis, tumor, trauma, earlier surgery). The slippage may be asymptomatic or may cause LBP, nerve root injury (the L5 root most frequently), symptomatic spinal stenosis, or CES in rare severe cases. A “step-off” on palpation or tenderness may be elicited near the segment that has “slipped” (most often L4 on L5 or occasionally L5 on S1). Focal anterolisthesis or retrolisthesis can occur at any cervical or lumbar level and be the source of neck or LBP. Plain x-rays of the low back or neck in flexion and extension will reveal movement at the abnormal spinal segment. Surgery is performed for spinal instability (slippage 5–8 mm) and considered for pain symptoms that do not respond to conservative measures (e.g., rest, physical therapy), cases with a progressive neurologic deficit, or scoliosis.

■ NEOPLASMS

Back pain is the most common neurologic symptom in patients with systemic cancer and is the presenting symptom in 20%. The cause is usually vertebral body metastasis (85–90%) but can also result from spread of cancer through the intervertebral foramen (especially with lymphoma), carcinomatous meningitis, or metastasis to the spinal cord. The thoracic spine is most often affected. Cancer-related back pain tends to be constant, dull, unrelieved by rest, and worse at night. By contrast, mechanical causes of LBP usually improve with rest. MRI, CT, and CT myelography are the studies of choice when spinal metastasis is suspected. Once a metastasis is found, imaging of the entire spine is essential, as it reveals additional tumor deposits in one-third of patients. MRI is preferred for soft tissue definition, but the most rapidly available imaging modality is best because the patient's condition may worsen quickly without intervention. Early diagnosis is crucial. A strong predictor of outcome is baseline neurologic function prior to diagnosis. Half to three-quarters of patients are nonambulatory at the time of diagnosis and few regain the ability to walk. **The management of spinal metastasis is discussed in detail in Chap. 90.**

■ INFECTIONS/INFLAMMATION

Vertebral osteomyelitis is most often caused by hematogenous seeding of staphylococci, but other bacteria or tuberculosis (Pott's disease) may be responsible. The primary source of infection is usually the skin or urinary tract. Other common sources of bacteremia are IV drug use, poor dentition, endocarditis, lung abscess, IV catheters, or postoperative wound sites. Back pain at rest, tenderness over the involved vertebra, and an elevated erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) are the most common findings in vertebral osteomyelitis. Fever or an elevated white blood cell count is found in a minority of patients. MRI and CT are sensitive and specific for early detection of osteomyelitis. The intervertebral disk can also be affected by infection (diskitis) and almost never by tumor. Extension of the infection posteriorly from the vertebral body can produce a spinal epidural abscess.

Spinal epidural abscess (Chap. 442) presents with back pain (aggravated by movement or palpation of the spinous process), fever, radiculopathy, or signs of spinal cord compression. The subacute development of two or more of these findings should increase suspicion for spinal epidural abscess. The abscess is best delineated by spine MRI and may track over multiple spinal levels.

Lumbar adhesive arachnoiditis with radiculopathy is due to fibrosis following inflammation within the subarachnoid space. The fibrosis results in nerve root adhesions and presents as back and leg pain associated with multifocal motor, sensory, or reflex changes. Causes of arachnoiditis include multiple lumbar operations (most common in the United States), chronic spinal infections (especially tuberculosis in the developing world), spinal cord injury, intrathecal hemorrhage, myelography (rare), intrathecal injections (glucocorticoids, anesthetics, or

other agents), and foreign bodies. The MRI shows clumped nerve roots on axial views or loculations of cerebrospinal fluid within the thecal sac. Clumped nerve roots should be distinguished from enlarged nerve roots seen with demyelinating polyneuropathy or neoplastic infiltration. Treatment is usually unsatisfactory. Microsurgical lysis of adhesions, dorsal rhizotomy, dorsal root ganglionectomy, and epidural glucocorticoids have been tried, but outcomes have been poor. Dorsal column stimulation for pain relief has produced varying results.

■ TRAUMA

A patient complaining of back pain and an inability to move the legs may have a spine fracture or dislocation; fractures above L1 place the spinal cord at risk for compression. Care must be taken to avoid further damage to the spinal cord or nerve roots by immobilizing the back or neck pending the results of radiologic studies. Vertebral fractures frequently occur in the absence of trauma in association with osteoporosis, glucocorticoid use, osteomyelitis, or neoplastic infiltration.

Sprains and Strains The terms *low back sprain*, *strain*, and *mechanically induced muscle spasm* refer to minor, self-limited injuries associated with lifting a heavy object, a fall, or a sudden deceleration such as in an automobile accident. These terms are used loosely and do not correlate with specific underlying pathologies. The pain is usually confined to the lower back. Patients with paraspinal muscle spasm often assume unusual postures.

Traumatic Vertebral Fractures Most traumatic fractures of the lumbar vertebral bodies result from injuries producing anterior wedging or compression. With severe trauma, the patient may sustain a fracture-dislocation or a “burst” fracture involving the vertebral body and posterior elements. Traumatic vertebral fractures are caused by falls from a height, sudden deceleration in an automobile accident, or direct injury. Neurologic impairment is common, and early surgical treatment is indicated. In victims of blunt trauma, CT scans of the chest, abdomen, or pelvis can be reformatted to detect associated vertebral fractures. Rules have been developed to avoid unnecessary spine imaging associated with low-risk trauma, but these studies typically exclude patients aged >65 years—a group that can sustain fractures with minor trauma.

■ METABOLIC CAUSES

Osteoporosis and Osteosclerosis Immobilization, osteomalacia, the postmenopausal state, renal disease, multiple myeloma, hyperparathyroidism, hyperthyroidism, metastatic carcinoma, or glucocorticoid use may accelerate osteoporosis and weaken the vertebral body, leading to compression fractures and pain. Up to two-thirds of compression fractures seen on radiologic imaging are asymptomatic. The most common nontraumatic vertebral body fractures are due to a postmenopausal cause, or to osteoporosis in adults >75 years old (Chap. 411). The risk of an additional vertebral fracture 1 year following a first vertebral fracture is 20%. The presence of fever, weight loss, fracture at a level above T4, any fracture in a young adult, or the predisposing conditions described above should increase suspicion for a cause other than typical osteoporosis. The sole manifestations of a compression fracture may be localized back or radicular pain exacerbated by movement and often reproduced by palpation over the spinous process of the affected vertebra.

Relief of acute pain can often be achieved with acetaminophen, NSAIDs, opioids, or a combination of these medications. Both pain and disability are improved with bracing. Antiresorptive drugs are not recommended in the setting of acute pain but are the preferred treatment to prevent additional fractures. Less than one-third of patients with prior compression fractures are adequately treated for osteoporosis despite the increased risk for future fractures; even fewer at-risk patients without a history of fracture are adequately treated. The literature for percutaneous vertebroplasty (PVP) or kyphoplasty for osteoporotic compression fractures associated with debilitating pain does not support their use.

Osteosclerosis, an abnormally increased bone density often due to Paget's disease, is readily identifiable on routine x-ray studies and can sometimes be a source of back pain. It may be associated with an isolated increase in alkaline phosphatase in an otherwise healthy older person. Spinal cord or nerve root compression can result from bony encroachment. The diagnosis of Paget's disease as the cause of a patient's back pain is a diagnosis of exclusion.

For further discussion of these bone disorders, see Chaps. 410, 411, and 412.

AUTOIMMUNE INFLAMMATORY ARTHRITIS

Autoimmune inflammatory disease of the spine can present with the insidious onset of low back, buttock, or neck pain. Examples include rheumatoid arthritis (RA) (Chap. 358), ankylosing spondylitis, reactive arthritis and psoriatic arthritis (Chap. 355), or inflammatory bowel disease (Chap. 326).

CONGENITAL ANOMALIES OF THE LUMBAR SPINE

Spondylosis is a bony defect in the vertebral pars interarticularis (a segment near the junction of the pedicle with the lamina), a finding present in up to 6% of adolescents. The cause is usually a stress microfracture in a congenitally abnormal segment. Multislice CT with multiplanar reformation is the most accurate modality for detecting spondylosis in adults. Symptoms may occur in the setting of a single injury, repeated minor injuries, or during a growth spurt. Spondylosis is the most common cause of persistent LBP in adolescents and is often associated with sports-related activities.

Scoliosis refers to an abnormal curvature in the coronal (lateral) plane of the spine. With *kyphoscoliosis*, there is, in addition, a forward curvature of the spine. The abnormal curvature may be congenital, due to abnormal spine development, acquired in adulthood due to degenerative spine disease, or progressive due to paraspinal neuromuscular disease. The deformity can progress until ambulation or pulmonary function is compromised.

Spina bifida occulta (closed spinal dysraphism) is a failure of closure of one or several vertebral arches posteriorly; the meninges and spinal cord are normal. A dimple or small lipoma may overlie the defect, but the skin is intact. Most cases are asymptomatic and discovered incidentally during a physical examination for back pain.

Tethered cord syndrome usually presents as a progressive cauda equina disorder (see below), although myelopathy may also be the initial manifestation. The patient is often a child or young adult who complains of perineal or perianal pain, sometimes following minor trauma. MRI studies typically reveal a low-lying conus (below L1 and L2) and a short and thickened filum terminale. The MRI findings also occur as incidental findings, sometimes during evaluation of unrelated LBP in adults.

REFERRED PAIN FROM VISCERAL DISEASE

Diseases of the thorax, abdomen, or pelvis may refer pain to the spinal segment that innervates the diseased organ. Occasionally, back pain may be the first and only manifestation. Upper abdominal diseases generally refer pain to the lower thoracic or upper lumbar region (eighth thoracic to the first and second lumbar vertebrae), lower abdominal diseases to the midlumbar region (second to fourth lumbar vertebrae), and pelvic diseases to the sacral region. Local signs (pain with spine palpation, paraspinal muscle spasm) are absent, and little or no pain accompanies routine movements.

Low Thoracic or Lumbar Pain with Abdominal Disease Tumors of the posterior wall of the stomach or duodenum typically produce epigastric pain (Chaps. 80 and 324), but back pain may occur if retroperitoneal extension is present. Fatty foods occasionally induce back pain associated with biliary or pancreatic disease. Pathology in retroperitoneal structures (hemorrhage, tumors, and pyelonephritis) can produce paraspinal pain that radiates to the lower abdomen, groin, or anterior thighs. A mass in the iliopsoas region can produce unilateral lumbar pain with radiation toward the groin, labia, or testicle. The sudden appearance of lumbar

pain in a patient receiving anticoagulants should prompt consideration of retroperitoneal hemorrhage.

Isolated LBP occurs in some patients with a contained rupture of an AAA. The classic clinical triad of abdominal pain, shock, and back pain occurs in <20% of patients. The diagnosis may be missed because the symptoms and signs can be nonspecific. Misdiagnoses include nonspecific back pain, diverticulitis, renal colic, sepsis, and myocardial infarction. A careful abdominal examination revealing a pulsatile mass (present in 50–75% of patients) is an important physical finding. Patients with suspected AAA should be evaluated with abdominal ultrasound, CT, or MRI (Chap. 280).

Sacral Pain with Gynecologic and Urologic Disease Pelvic organs rarely cause isolated LBP. Uterine malposition (retroversion, descensus, and prolapse) may cause traction on the uterosacral ligament. The pain is referred to the sacral region, sometimes appearing after prolonged standing. Endometriosis or uterine cancers can invade the uterosacral ligaments. Pain associated with endometriosis is typically premenstrual and often continues until it merges with menstrual pain.

Menstrual pain with poorly localized, cramping pain can radiate down the legs. LBP that radiates into one or both thighs is common in the last weeks of pregnancy. Continuous and worsening pain unrelieved by rest or at night may be due to neoplastic infiltration of nerves or nerve roots.

Urologic sources of lumbosacral back pain include chronic prostatitis, prostate cancer with spinal metastasis (Chap. 87), and diseases of the kidney or ureter. Infectious, inflammatory, or neoplastic renal diseases may produce ipsilateral lumbosacral pain, as can renal artery or vein thrombosis. Paraspinal lumbar pain may be a symptom of ureteral obstruction due to nephrolithiasis.

OTHER CAUSES OF BACK PAIN

Postural Back Pain There is a group of patients with nonspecific chronic low back pain (CLBP) in whom no specific anatomic lesion can be found despite exhaustive investigation. Exercises to strengthen the paraspinal and abdominal muscles are sometimes helpful. CLBP may be encountered in patients who seek financial compensation; in malingerers; or in those with concurrent substance abuse. Many patients with CLBP have a history of psychiatric illness (depression, anxiety states) or childhood trauma (physical or sexual abuse) that antedates the onset of back pain. Preoperative psychological assessment has been used to exclude patients with marked psychological impairments that predict a poor surgical outcome from spine surgery.

Idiopathic The cause of LBP occasionally remains unclear. Some patients have had multiple operations for disk disease. The original indications for surgery may have been questionable, with back pain only, no definite neurologic signs, or a minor disk bulge noted on CT or MRI. Scoring systems based on neurologic signs, psychological factors, physiologic studies, and imaging studies have been devised to minimize the likelihood of unsuccessful surgery.

GLOBAL CONSIDERATIONS

While many of the history and examination features described in this chapter apply to all patients, information regarding the global epidemiology and prevalence of LBP is limited. The Global Burden of Diseases Study 2019 reported that LBP represented the #1 cause overall for total years lived with disability (YLD), and #9 overall as a cause of disability-related life years (DALYs). These numbers increased substantially from 1990 estimates, and with the aging of the population worldwide, the numbers of individuals suffering from LBP are expected to increase further in the future. Although rankings for LBP generally were higher in developed regions, a high burden exists in every part of the world. An area of uncertainty is the degree to which regional differences exist in terms of the specific etiologies of LBP and how these are managed. For example, the most common cause of arachnoiditis in developing countries is a prior spinal infection, but in developed countries the most frequent cause is multiple lumbar spine surgeries.

TREATMENT

Back Pain

Management is considered separately for acute and chronic low back pain syndromes without radiculopathy, and for back pain with radiculopathy.

ACUTE LOW BACK PAIN WITHOUT RADICULOPATHY

This is defined as pain of <12 weeks duration. Full recovery can be expected in >85% of adults with ALBP without leg pain. Most have purely “mechanical” symptoms (i.e., pain that is aggravated by motion and relieved by rest).

The initial assessment is focused on excluding serious causes of spine pathology that require urgent intervention, including infection, cancer, or trauma. Risk factors for a serious cause of ALBP are shown in Table 17-1. Laboratory and imaging studies are unnecessary if risk factors are absent. CT, MRI, or plain spine films are rarely indicated in the first month of symptoms unless a spine fracture, tumor, or infection is suspected.

The prognosis of ALBP is generally excellent; however, episodes tend to recur, and as many as two-thirds of patients will experience a second episode within 1 year. Most patients do not seek medical care and improve on their own. Even among those seen in primary care, two-thirds report substantial improvement after 7 weeks. This high likelihood of spontaneous improvement can mislead clinicians and patients about the efficacy of treatment interventions, highlighting the importance of rigorous prospective trials. Many treatments commonly used in the past are now known to be ineffective, including bed rest and lumbar traction.

Clinicians should reassure and educate patients that improvement is very likely and instruct them in self-care. Satisfaction and the likelihood of follow-up increase when patients are educated about prognosis, evidence-based treatments, appropriate activity modifications, and strategies to prevent future exacerbations. Counseling patients about the risks of overtreatment is another important part of the discussion. Patients who report that they did not receive an adequate explanation for their symptoms are likely to request further diagnostic tests.

In general, bed rest should be avoided for relief of severe symptoms or limited to a day or two at most. Several randomized trials suggest that bed rest does not hasten the pace of recovery. In general, early resumption of normal daily physical activity should be encouraged, avoiding only strenuous manual labor. Advantages of early ambulation for ALBP also include maintenance of cardiovascular conditioning; improved bone, cartilage, and muscle strength; and increased endorphin levels. Specific back exercises or early vigorous exercise have not shown benefits for acute back pain. Empiric use of heating pads or blankets is sometimes helpful.

NSAIDs and Acetaminophen Evidence-based guidelines recommend over-the-counter medicines such as NSAIDs and acetaminophen as first-line options for treatment of ALBP. In otherwise healthy patients, a trial of NSAIDs can be followed by acetaminophen for time-limited periods. In theory, the anti-inflammatory effects of NSAIDs might provide an advantage over acetaminophen to suppress inflammation that accompanies many causes of ALBP, but in practice there is no clinical evidence to support the superiority of NSAIDs. The risk of renal and gastrointestinal toxicity with NSAIDs is increased in patients with preexisting medical comorbidities (e.g., renal insufficiency, cirrhosis, prior gastrointestinal hemorrhage, use of anticoagulants or glucocorticoids, heart failure). Some patients elect to take acetaminophen and an NSAID together in hopes of a more rapid benefit.

Muscle Relaxants Skeletal muscle relaxants, such as cyclobenzaprine or methocarbamol, may be useful, but sedation is a common side effect. Limiting the use of muscle relaxants to nighttime only may be an option for patients with back pain that interferes with sleep.

Opioids There is no good evidence to support the use of opioid analgesics or tramadol as first-line therapy for ALBP. Their use is best reserved for patients who cannot tolerate acetaminophen or NSAIDs and for those with severe refractory pain. Also, the duration of opioid treatment for ALBP should be strictly limited to 3–7 days. As with muscle relaxants, these drugs are often sedating, so it may be useful to prescribe them at nighttime only. Side effects of short-term opioid use include nausea, constipation, and pruritus; risks of long-term opioid use include hypersensitivity to pain, hypogonadism, and dependency. Falls, fractures, driving accidents, and fecal impaction are other risks. The clinical efficacy of opioids for chronic pain beyond 16 weeks of use is unproven.

Mounting evidence of morbidity from long-term opioid therapy (including overdose, dependency, addiction, falls, fractures, accident risk, and sexual dysfunction) has prompted efforts to reduce its use for chronic pain, including back pain (Chap. 13). When used, safety may be improved with automated notices for high doses, early refills, prescriptions from multiple pharmacies, overlapping opioid and benzodiazepine prescriptions, and in the United States by state-based prescription drug monitoring programs (PDMPs). A recent study indicated that most patients with opioid use disorder presenting to emergency departments had no prescriptions recorded in the PDMP, reflecting other methods used to obtain opioids. Greater access to alternative treatments for chronic pain, such as tailored exercise programs and cognitive behavioral therapy (CBT), may also reduce opioid prescribing.

Other Approaches There is no evidence to support use of oral or injected glucocorticoids, antiepileptics, antidepressants, or therapies for neuropathic pain such as gabapentin or herbal therapies. Commonly used nonpharmacologic treatments for ALBP are also of unproven benefit, including spinal manipulation, physical therapy, massage, acupuncture, laser therapy, therapeutic ultrasound, corsets, transcutaneous electrical nerve stimulation (TENS), special mattresses, or lumbar traction. Although important for chronic pain, use of back exercises for ALBP are generally not supported by clinical evidence. There is no convincing evidence regarding the value of ice or heat applications for ALBP; however, many patients report temporary symptomatic relief from ice or frozen gel packs just before sleep, and heat may produce a short-term reduction in pain after the first week. Patients often report improved satisfaction with the care that they receive when they actively participate in the selection of symptomatic approaches.

CHRONIC LOW BACK PAIN WITHOUT RADICULOPATHY

Back pain is considered chronic when the symptoms last >12 weeks; it accounts for 50% of total back pain costs. Risk factors include obesity, female gender, older age, prior history of back pain, restricted spinal mobility, pain radiating into a leg, high levels of psychological distress, poor self-rated health, minimal physical activity, smoking, job dissatisfaction, and widespread pain. In general, the same treatments that are recommended for ALBP can be useful for patients with CLBP. In this setting, however, the benefit of opioid therapy or muscle relaxants is less clear. In general, improved activity tolerance is the primary goal, while pain relief is secondary.

Some observers have raised concerns that CLBP may often be overtreated. For CLBP without radiculopathy, multiple guidelines explicitly recommend against use of SSRIs, any type of injection, TENS, lumbar supports, traction, radiofrequency facet joint denervation, intradiskal electrothermal therapy, or intradiskal radiofrequency thermocoagulation. On the other hand, exercise therapy and treatment of depression appear to be useful and underused.

Exercise Programs Evidence supports the use of exercise therapy to alleviate pain symptoms and improve function. Exercise can be one of the mainstays of treatment for CLBP. Effective regimens have generally included a combination of core-strengthening exercises, stretching, and gradually increasing aerobic exercise. A program of supervised exercise can improve compliance. Supervised intensive

physical exercise or “work hardening” regimens have been effective in returning some patients to work, improving walking distance, and reducing pain. In addition, some forms of yoga have been evaluated in randomized trials and may be helpful for patients who are interested.

Intensive multidisciplinary rehabilitation programs can include daily or frequent physical therapy, exercise, CBT, a workplace evaluation, and other interventions. For patients who have not responded to other approaches, such programs appear to offer some benefit. Systematic reviews, however, suggest that the evidence and benefits are limited.

Nonopioid Medications Medications for CLBP may include short courses of NSAIDs or acetaminophen. Duloxetine is approved for the treatment of CLBP (60 mg daily) and may also treat coincident depression. Tricyclic antidepressants can provide modest pain relief for some patients without evidence of depression. Depression is common among patients with chronic pain and should be appropriately treated.

Cognitive Behavioral Therapy CBT is based on evidence that psychological and social factors, as well as somatic pathology, are important in the genesis of chronic pain and disability; CBT focuses on efforts to identify and modify patients’ thinking about their condition. In one randomized trial, CBT reduced disability and pain in patients with CLBP. Such behavioral treatments appear to provide benefits similar in magnitude to exercise therapy.

Complementary Medicine Back pain is the most frequent reason for seeking complementary and alternative treatments. Spinal manipulation or massage therapy may provide short-term relief, but long-term benefit is unproven. Biofeedback has not been studied rigorously. There is no convincing evidence that either TENS, laser therapy, or ultrasound are effective in treating CLBP. Rigorous trials of acupuncture suggest that true acupuncture is not superior to sham acupuncture, but that both may offer an advantage over routine care. Whether this is due entirely to placebo effects provided even by sham acupuncture is uncertain.

Injections and Other Interventions Various injections, including epidural glucocorticoid injections, facet joint injections, and trigger point injections, have been used for treating CLBP. However, in the absence of radiculopathy, there is no clear evidence that these approaches are sustainably effective.

Injection studies are sometimes used diagnostically to help determine the anatomic source of back pain. Pain relief following a glucocorticoid and anesthetic injection into a facet or medial branch block are used as evidence that the facet joint is the pain source; however, the possibility that the response was a placebo effect or due to systemic absorption of the glucocorticoids is difficult to exclude.

Another category of intervention for CLBP is electrothermal and radiofrequency therapy. Intradiskal therapy has been proposed using energy to thermocoagulate and destroy nerves in the intervertebral disk, using specially designed catheters or electrodes. Current evidence does not support the use of discography to identify a specific disk as the pain source, or the use of intradiskal electrothermal or radiofrequency therapy for CLBP.

Radiofrequency denervation is sometimes used to destroy nerves that are thought to mediate pain, and this technique has been used for facet joint pain (with the target nerve being the medial branch of the primary dorsal ramus), for back pain thought to arise from the intervertebral disk (ramus communicans), and radicular back pain (dorsal root ganglia). These interventional therapies have not been studied in sufficient detail to draw firm conclusions regarding their value for CLBP.

Surgery Surgical intervention for CLBP without radiculopathy has been evaluated in a number of randomized trials. The case for fusion surgery for CLBP without radiculopathy is weak. While some studies have shown modest benefit, there has been no benefit when compared to an active medical treatment arm, often including highly structured, rigorous rehabilitation combined with CBT. The

use of bone matrix protein (BMP) instead of iliac crest graft for the fusion was shown to increase hospital costs and length of stay but not improve clinical outcomes.

Guidelines suggest that referral for an opinion on spinal fusion can be considered for patients who have completed an optimal nonsurgical treatment program (including combined physical and psychological treatment) and who have persistent severe back pain for which they would consider surgery. The high cost, wide geographic variations, and rapidly increasing rates of spinal fusion surgery have prompted scrutiny regarding the lack of standardization of appropriate indications. Some insurance carriers have begun to limit coverage for the most controversial indications, such as LBP without radiculopathy.

Lumbar disk replacement with prosthetic disks is US Food and Drug Administration-approved for uncomplicated patients needing single-level surgery at the L3-S1 levels. The disks are generally designed as metal plates with a polyethylene cushion sandwiched in between. The trials that led to approval of these devices were not blinded. When compared to spinal fusion, the artificial disks were “not inferior.” Long-term follow-up is needed to determine device failure rates over time. Serious complications are somewhat more likely with the artificial disk. This treatment remains controversial for CLBP.

LOW BACK PAIN WITH RADICULOPATHY

A common cause of back pain with radiculopathy is a herniated disk affecting the nerve root and producing back pain with radiation down the leg. The term *sciatica* is used when the leg pain radiates posteriorly in a sciatic or L5/S1 distribution. The prognosis for acute low back and leg pain with radiculopathy due to disk herniation is generally favorable, with most patients showing substantial improvement over months. Serial imaging studies suggest spontaneous regression of the herniated portion of the disk in two-thirds of patients over 6 months. Nonetheless, several important treatment options provide symptomatic relief while the healing process unfolds.

Resumption of normal activity is recommended. Randomized trial evidence suggests that bed rest is ineffective for treating sciatica as well as back pain alone. Acetaminophen and NSAIDs are useful for pain relief, although severe pain may require short courses (3–7 days) of opioid analgesics. Opioids are superior for acute pain relief in the emergency department.

Epidural glucocorticoid injections have a role in providing symptom relief for acute lumbar radiculopathy due to a herniated disk, but do not reduce the use of subsequent surgical intervention. A brief course of high-dose oral glucocorticoids (methylprednisolone dose pack) for 3 days followed by a rapid taper over 4 more days can be helpful for some patients with acute disk-related radiculopathy, although this specific regimen has not been studied rigorously.

Diagnostic nerve root blocks have been advocated to determine if pain originates from a specific nerve root. However, improvement may result even when the nerve root is not responsible for the pain; this may occur as a placebo effect, from a pain-generating lesion located distally along the peripheral nerve, or from effects of systemic absorption.

Urgent surgery is recommended for patients who have evidence of CES or spinal cord compression, generally manifesting as combinations of bowel or bladder dysfunction, diminished sensation in a saddle distribution, a sensory level on the trunk, and bilateral leg weakness or spasticity. Surgical intervention is also indicated for patients with progressive motor weakness due to nerve root injury demonstrated on clinical examination or EMG.

Surgery is also an important option for patients who have disabling radicular pain despite optimal conservative treatment. Because patients with a herniated disk and sciatica generally experience rapid improvement over weeks, most experts do not recommend considering surgery unless the patient has failed to respond to a minimum of 6–8 weeks of nonsurgical management. For patients who have not improved, randomized trials show that surgery results in more rapid pain relief than nonsurgical treatment. However, after

2 years of follow-up, patients appear to have similar pain relief and functional improvement with or without surgery. Thus, both treatment approaches are reasonable, and patient preferences and needs (e.g., rapid return to employment) strongly influence decision-making. Some patients will want the fastest possible relief and find surgical risks acceptable. Others will be more risk-averse and more tolerant of symptoms and will choose watchful waiting, especially if they understand that improvement is likely in the end.

The usual surgical procedure is a partial hemilaminectomy with excision of the prolapsed disk (discectomy). Minimally invasive techniques have gained in popularity in recent years, but some evidence suggests they may be less effective than standard surgical techniques, with more residual back pain, leg pain, and higher rates of rehospitalization. Fusion of the involved lumbar segments should be considered only if significant spinal instability is present (i.e., degenerative spondylolisthesis). The costs associated with lumbar interbody fusion have increased dramatically in recent years. There are no large prospective, randomized trials comparing fusion to other types of surgical intervention. In one study, patients with persistent LBP despite an initial discectomy fared no better with spine fusion than with a conservative regimen of cognitive intervention and exercise. Artificial disks, as discussed above, are used in Europe; their utility remains controversial in the United States.

PAIN IN THE NECK AND SHOULDER

Neck pain, which usually arises from diseases of the cervical spine and soft tissues of the neck, is common, typically precipitated by movement, and may be accompanied by focal tenderness and limitation of motion. Many of the earlier comments made regarding causes of LBP also apply to disorders of the cervical spine. The text below will emphasize differences. Pain arising from the brachial plexus, shoulder, or peripheral nerves can be confused with cervical spine disease (**Table 17-4**), but the history and examination usually identify a more distal origin for the pain. When the site of nerve tissue injury is unclear, EMG studies can localize the lesion. Cervical spine trauma, disk disease, or spondylosis with intervertebral foraminal narrowing may be asymptomatic or painful and can produce a myelopathy, radiculopathy, or both. The same risk factors for serious causes of LBP also apply to neck pain with the additional feature that neurologic signs of myelopathy (incontinence, sensory level, spastic legs) may also occur. Lhermitte's sign, an electrical shock down the spine with neck flexion, suggests involvement of the cervical spinal cord.

■ TRAUMA TO THE CERVICAL SPINE

Trauma (fractures, subluxation) places the spinal cord at risk for compression. Motor vehicle accidents, violent crimes, or falls account for 87% of cervical spinal cord injuries (**Chap. 442**). Immediate immobilization of the neck is essential to minimize further spinal cord injury from movement of unstable cervical spine segments. A CT scan is the diagnostic procedure of choice for detection of acute fractures following severe trauma; plain x-rays are used for lesser degrees of trauma or in settings where CT is unavailable. When traumatic injury to the vertebral arteries or cervical spinal cord is suspected, visualization by MRI with magnetic resonance angiography is preferred.

The decision to obtain imaging should be based on the clinical context of the injury. The National Emergency X-Radiography Utilization Study (NEXUS) low-risk criteria established that normally alert patients without palpable tenderness in the midline; intoxication; neurologic deficits; or painful distracting injuries were very unlikely to have sustained a clinically significant traumatic injury to the cervical spine. The Canadian C-spine rule recommends that imaging should be obtained following neck region trauma if the patient is >65 years old or has limb paresthesias or if there was a dangerous mechanism for the injury (e.g., bicycle collision with tree or parked car, fall from height >3 ft or five stairs, diving accident). These guidelines are helpful but must be tailored to individual circumstances; for example, patients with advanced osteoporosis, glucocorticoid use, or cancer may warrant imaging after even mild trauma.

Whiplash injury is due to rapid flexion and extension of the neck, usually from automobile accidents. The likely mechanism involves injury to the facet joints. This diagnosis should not be applied to patients with fractures, disk herniation, head injury, focal neurologic findings, or altered consciousness. Up to 50% of persons reporting whiplash injury acutely have persistent neck pain 1 year later. When personal compensation for pain and suffering was removed from the Australian health care system, the prognosis for recovery at 1 year improved. Imaging of the cervical spine is not cost-effective acutely but is useful to detect disk herniations when symptoms persist for >6 weeks following the injury. Severe initial symptoms have been associated with a poor long-term outcome.

■ CERVICAL DISK DISEASE

Degenerative cervical disk disease is very common and usually asymptomatic. Herniation of a lower cervical disk is a common cause of pain or tingling in the neck, shoulder, arm, or hand. Neck pain, stiffness, and a range of motion limited by pain are the usual manifestations.

TABLE 17-4 Cervical Radiculopathy: Neurologic Features

CERVICAL NERVE ROOT	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
C5	Biceps	Lateral deltoid	Rhombooids ^a (elbow extends backward with hand on hip) Infraspinatus ^a (arm rotates externally with elbow flexed at the side) Deltoid ^a (arm raised laterally 30°–45° from the side)	Lateral arm, medial scapula
C6	Biceps	Palmar thumb/index finger Dorsal hand/lateral forearm	Biceps ^a (arm flexed at the elbow in supination) Pronator teres (forearm pronated)	Lateral forearm, thumb/index fingers
C7	Triceps	Middle finger Dorsal forearm	Triceps ^a (forearm extension, flexed at elbow) Wrist/finger extensors ^a	Posterior arm, dorsal forearm, dorsal hand
C8	Finger flexors	Palmar surface of little finger Medial hand and forearm	Abductor pollicis brevis (abduction of thumb) First dorsal interosseous (abduction of index finger) Abductor digiti minimi (abduction of little finger)	Fourth and fifth fingers, medial hand and forearm
T1	Finger flexors	Axilla, medial arm, anteromedial forearm	Abductor pollicis brevis (abduction of thumb) First dorsal interosseous (abduction of index finger) Abductor digiti minimi (abduction of little finger)	Medial arm, axilla

^aThese muscles receive the majority of innervation from this root.

Herniated cervical disks are responsible for ~25% of cervical radiculopathies. Extension and lateral rotation of the neck narrow the ipsilateral intervertebral foramen and may reproduce radicular symptoms (Spurling's sign). In young adults, acute nerve root compression from a ruptured cervical disk is often due to trauma. Cervical disk herniations are usually posterolateral near the lateral recess. Typical patterns of reflex, sensory, and motor changes that accompany cervical nerve root lesions are summarized in Table 17-4. Although the classic patterns are clinically helpful, there are numerous exceptions because (1) there is overlap in sensory function between adjacent nerve roots, (2) symptoms and signs may be evident in only part of the injured nerve root territory, and (3) the location of pain is the most variable of the clinical features.

CERVICAL SPONDYLOYSIS

Osteoarthritis of the cervical spine may produce neck pain that radiates into the back of the head, shoulders, or arms, or may be the source of headaches in the posterior occipital region (supplied by the C2-C4 nerve roots). Osteophytes, disk protrusions, or hypertrophic facet or uncovertebral joints may alone or in combination compress one or several nerve roots at the intervertebral foramina; these causes together account for 75% of cervical radiculopathies. The roots most commonly affected are C7 and C6. Narrowing of the spinal canal by osteophytes, ossification of the posterior longitudinal ligament (OPLL), or a large central disk may compress the cervical spinal cord and produce signs of myelopathy alone or radiculopathy with myelopathy (myeloradiculopathy). When little or no neck pain accompanies cervical cord involvement, other diagnoses to be considered include amyotrophic lateral sclerosis (Chap. 437), multiple sclerosis (Chap. 444), spinal cord tumors, or syringomyelia (Chap. 442). Cervical spondylotic myelopathy should be considered even when the patient presents with symptoms or spinal cord signs in the legs only. MRI is the study of choice to define soft tissues in the cervical region including the spinal cord, whereas plain CT is optimal to identify bone pathology including foraminal, lateral recess, OPLL, or spinal canal stenosis. In spondylotic myelopathy, focal enhancement by MRI, sometimes in a characteristic "pancake pattern," may be present at the site of maximal cord compression.

There is no evidence to support prophylactic surgery for asymptomatic cervical spinal stenosis unaccompanied by myelopathic signs or abnormal spinal cord findings on MRI, except in the setting of *dynamic instability* (see spondylolisthesis above). If the patient has postural neck pain, a prior history of whiplash or other spine/head injury, a Lhermitte sign, or preexisting listhesis at the stenotic segment on cervical MRI or CT, then cervical spine flexion-extension x-rays or MRI are indicated to look for dynamic instability. Surgical intervention is not recommended for patients with listhesis alone, unaccompanied by dynamic instability.

OTHER CAUSES OF NECK PAIN

Rheumatoid arthritis (RA) (Chap. 358) of the cervical facet joints produces neck pain, stiffness, and limitation of motion. Synovitis of the atlantoaxial joint (C1-C2; Fig. 17-2) may damage the transverse ligament of the atlas, producing forward displacement of the atlas on the axis (atlantoaxial subluxation). Radiologic evidence of atlantoaxial subluxation occurs in up to 30% of patients with RA and plain x-ray films of the neck should be routinely performed preoperatively to assess the risk of neck hyperextension in patients requiring intubation. The degree of subluxation correlates with the severity of erosive disease. When subluxation is present, careful assessment is important to identify early signs of myelopathy that could be a harbinger of life-threatening spinal cord compression. Surgery should be considered when myelopathy or spinal instability is present. *Ankylosing spondylitis* is another cause of neck pain and less commonly atlantoaxial subluxation.

Acute *herpes zoster* can present as acute posterior occipital or neck pain prior to the outbreak of vesicles. *Neoplasms* metastatic to the cervical spine, *infections* (osteomyelitis and epidural abscess), and *metabolic bone diseases* may be the cause of neck pain, as discussed

above. Neck pain may also be referred from the heart with coronary artery ischemia (cervical angina syndrome). Rheumatologic disease should be considered if the neck pain is accompanied by shoulder or hip girdle pain.

THORACIC OUTLET SYNDROMES

The thoracic outlet contains the first rib, the subclavian artery and vein, the brachial plexus, the clavicle, and the lung apex. Injury to these structures may result in postural or movement-induced pain around the shoulder and supraclavicular region, classified as follows.

True neurogenic thoracic outlet syndrome (TOS) is an uncommon disorder resulting from compression of the lower trunk of the brachial plexus or ventral rami of the C8 or T1 nerve roots, caused most often by an anomalous band of cartilaginous tissue connecting an elongate transverse process at C7 with the first rib. Pain is mild or may be absent. Signs include weakness and wasting of intrinsic muscles of the hand and diminished sensation on the palmar aspect of the fifth digit. An anteroposterior cervical spine x-ray will show an elongate C7 transverse process (an anatomic marker for the anomalous cartilaginous band), and EMG and NCSs confirm the diagnosis. Treatment consists of surgical resection of the anomalous band. The weakness and wasting of intrinsic hand muscles typically do not improve, but surgery halts the insidious progression of weakness.

Arterial TOS results from compression of the subclavian artery by a cervical rib, resulting in poststenotic dilatation of the artery and in some cases secondary thrombus formation. Blood pressure is reduced in the affected limb, and signs of emboli may be present in the hand. Neurologic signs are absent. Ultrasound can confirm the diagnosis noninvasively. Treatment is with thrombolysis or anticoagulation (with or without embolectomy) and surgical excision of the cervical rib compressing the subclavian artery.

Venous TOS is due to subclavian vein thrombosis resulting in swelling of the arm and pain. The vein may be compressed by a cervical rib or anomalous scalene muscle. Venography is the diagnostic test of choice.

Disputed TOS accounts for 95% of patients diagnosed with TOS; chronic arm and shoulder pain are prominent and of unclear cause. The lack of sensitive and specific findings on physical examination or specific markers for this condition results in diagnostic uncertainty. The role of surgery in disputed TOS is controversial. Major depression, chronic symptoms, work-related injury, and diffuse arm symptoms predict poor surgical outcomes. Multidisciplinary pain management is a conservative approach, although treatment is often unsuccessful.

BRACHIAL PLEXUS AND NERVES

Pain from injury to the brachial plexus or peripheral nerves of the arm can occasionally mimic referred pain of cervical spine origin, including cervical radiculopathy, but the pain typically begins distal to the posterior neck region in the shoulder girdle or upper arm. Neoplastic infiltration of the lower trunk of the brachial plexus may produce shoulder or supraclavicular pain radiating down the arm, numbness of the fourth and fifth fingers or medial forearm, and weakness of intrinsic hand muscles innervated by the lower trunk and medial cord of the brachial plexus. Delayed radiation injury may produce weakness in the upper arm or numbness of the lateral forearm or arm due to involvement of the upper trunk and lateral cord of the plexus. Pain is less common and less severe than with neoplastic infiltration. A Pancoast tumor of the lung (Chap. 78) is another cause and should be considered, especially when a concurrent Horner's syndrome is present. *Acute brachial neuritis* is often confused with radiculopathy; the acute onset of severe shoulder or scapular pain is followed typically over days by weakness of the proximal arm and shoulder girdle muscles innervated by the upper brachial plexus. The onset may be preceded by an infection, vaccination, or minor surgical procedure. The long thoracic nerve may be affected, resulting in a winged scapula. Brachial neuritis may also present as an isolated paralysis of the diaphragm with or without involvement of other nerves of the upper limb. Recovery may take up to 3 years, and full functional recovery can be expected in the majority of patients.

Occasional cases of carpal tunnel syndrome produce pain and paresthesias extending into the forearm, arm, and shoulder resembling a C5 or C6 root lesion. Lesions of the radial or ulnar nerve can also mimic radiculopathy, at C7 or C8, respectively. EMG and NCSs can accurately localize lesions to the nerve roots, brachial plexus, or peripheral nerves.

For further discussion of peripheral nerve disorders, see Chap. 446.

■ SHOULDER

Pain arising from the shoulder can on occasion mimic pain from the spine. If symptoms and signs of radiculopathy are absent, then the differential diagnosis includes mechanical shoulder pain (bicipital tendonitis, frozen shoulder, bursitis, rotator cuff tear, dislocation, adhesive capsulitis, or rotator cuff impingement under the acromion) and referred pain (subdiaphragmatic irritation, angina, Pancoast tumor). Mechanical pain is often worse at night, associated with local shoulder tenderness and aggravated by passive abduction, internal rotation, or extension of the arm. Demonstrating normal passive full range of motion of the arm at the shoulder without worsening the usual pain can help exclude mechanical shoulder pathology as a cause of neck region pain. Pain from shoulder disease may radiate into the arm or hand, but focal neurologic signs (sensory, motor, or reflex changes) are absent.

■ GLOBAL CONSIDERATIONS

Many of the considerations described above for LBP also apply to neck pain. The Global Burden of Diseases Study 2019 reported that neck pain ranked second only to back pain as a cause of total years lived with disability (YLD). In general, neck pain rankings were also higher in developed regions of the world.

TREATMENT

Neck Pain Without Radiculopathy

The evidence regarding treatment for neck pain is less comprehensive than that for LBP, but the approach is remarkably similar in many respects. As with LBP, spontaneous improvement is the norm for acute neck pain. The usual goals of therapy are to promote a rapid return to normal function and provide pain relief while healing proceeds.

Acute neck pain is often treated with NSAIDs, acetaminophen, cold packs, or heat, alone or in combination while awaiting recovery. Patients should be specifically educated regarding the favorable natural history of acute neck pain to avoid unrealistic fear and inappropriate requests for imaging and other tests. For patients kept awake by symptoms, cyclobenzaprine (5–10 mg) at night can help relieve muscle spasm and promote drowsiness. For patients with neck pain unassociated with trauma, supervised exercise with or without mobilization appears to be effective. Exercises often include shoulder rolls and neck stretches. The evidence in support of non-surgical treatments for whiplash-associated disorders is generally of limited quality and neither supports nor refutes the common treatments used for symptom relief. Gentle mobilization of the cervical spine combined with exercise programs may be beneficial. Evidence is insufficient to recommend the use of cervical traction, TENS, ultrasound, trigger point injections, botulinum toxin injections, tricyclic antidepressants, and SSRIs for acute or chronic neck pain. Some patients obtain modest pain relief using a soft neck collar; there is little risk or cost. Massage can produce temporary pain relief.

For patients with chronic neck pain, supervised exercise programs can provide symptom relief and improve function. Acupuncture provided short-term benefit for some patients when compared to a sham procedure and is an option. Spinal manipulation alone has not been shown to be effective and carries a risk for injury. Surgical treatment for chronic neck pain without radiculopathy or spine instability is not recommended.

Neck Pain With Radiculopathy

The natural history of acute neck pain with radiculopathy due to disk disease is favorable, and many patients will improve without specific therapy. Although there are no randomized trials of NSAIDs for neck pain, a course of NSAIDs, acetaminophen, or both, with or without muscle relaxants, and avoidance of activities that trigger symptoms are reasonable as initial therapy. Gentle supervised exercise and avoidance of inactivity are reasonable as well. A short course of high-dose oral glucocorticoids with a rapid taper, or epidural steroids administered under imaging guidance can be effective for acute or subacute disk-related cervical radicular pain, but have not been subjected to rigorous trials. The risk of injection-related complications is higher in the neck than the low back; vertebral artery dissection, dural puncture, spinal cord injury, and embolism in the vertebral arteries have all been reported. Opioid analgesics can be used in the emergency department and for short courses as an outpatient. Soft cervical collars can be modestly helpful by limiting spontaneous and reflex neck movements that exacerbate pain; hard collars are in general poorly tolerated.

If cervical radiculopathy is due to bony compression from cervical spondylosis with foraminal narrowing, periodic follow-up to assess for progression is indicated and consideration of surgical decompression is reasonable. Surgical treatment can produce rapid pain relief, although it is unclear if long-term functional outcomes are improved over nonsurgical therapy. Indications for cervical disk surgery include a progressive motor deficit due to nerve root compression, functionally limiting pain that fails to respond to conservative management, or spinal cord compression. In other circumstances, clinical improvement over time regardless of therapeutic intervention is common.

Surgical treatments include anterior cervical diskectomy alone, laminectomy with diskectomy, or diskectomy with fusion. The risk of subsequent radiculopathy or myelopathy at cervical segments adjacent to a fusion is ~3% per year and 26% per decade. Although this risk is sometimes portrayed as a late complication of surgery, it may also reflect the natural history of degenerative cervical disk disease.

■ FURTHER READING

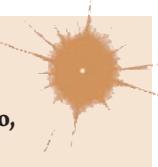
- AGENCY FOR HEALTHCARE RESEARCH AND QUALITY (AHRQ): Non-invasive treatments for low back pain. AHRQ Publication No. 16-EHC004-EF. February 2016, <https://effectivehealthcare.ahrq.gov/ehc/products/553/2178/back-pain-treatment-report-160229.pdf>
- AUSTEVOLL IM et al: Decompression with or without fusion in degenerative lumbar spondylolisthesis. *N Engl J Med* 385:526, 2021.
- BAILEY CS et al: Surgery versus conservative care for persistent sciatica lasting 4 to 12 months. *N Engl J Med* 19;382:1093, 2020.
- CIEZA A et al: Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 396:2006, 2021.
- ENGSTROM JW: Physical and Neurologic Examination. In Steinmetz et al (eds). *Benzel's Spine Surgery*, 5th ed. Philadelphia, Elsevier, 2021.
- GOLDBERG H et al: Oral steroids for acute radiculopathy due to a herniated lumbar disk. *JAMA* 313:1915, 2015.
- HAWK K et al: Past-year prescription drug monitoring program opioid prescriptions and self-reported opioid use in an emergency department population with opioid use disorder. *Acad Emerg Med* 25:508, 2018.
- JARVIK JG et al: Association of early imaging for back pain with clinical outcomes in older adults. *JAMA* 313:1143, 2015.
- KATZ JN, HARRIS MB: Clinical practice. Lumbar spinal stenosis. *N Engl J Med* 358:818, 2008.
- THEODORE N: Degenerative cervical spondylosis. *N Engl J Med* 383:159, 2020.
- ZYGOURAKIS CC et al: Geographic and hospital variation in cost of lumbar laminectomy and lumbar fusion for degenerative conditions. *Neurosurgery* 81:331, 2017.

Section 2 Alterations in Body Temperature

18

Fever

Neeraj K. Surana, Charles A. Dinarello,
Reuven Porat



Body temperature is controlled by the hypothalamus. Neurons in both the preoptic anterior hypothalamus and the posterior hypothalamus receive two kinds of signals: one from peripheral nerves that transmit information from warmth/cold receptors in the skin and the other from the temperature of the blood bathing the region. These two types of signals are integrated by the thermoregulatory center of the hypothalamus to maintain normal temperature. In a neutral temperature environment, the human metabolic rate produces more heat than is necessary to maintain the core body temperature in the range of 36.5–37.5°C (97.7–99.5°F).

A normal body temperature is ordinarily maintained despite environmental variations because the hypothalamic thermoregulatory center balances the excess heat production derived from metabolic activity in muscle and the liver with heat dissipation from the skin and lungs. According to a study of >35,000 individuals ≥18 years of age seen in routine medical visits, the mean oral temperature is 36.6°C (95% confidence interval, 35.7–37.3°C). In light of this study, *a temperature of >37.7°C (>99.9°F), which represents the 99th percentile for healthy individuals, defines a fever.* Importantly, higher ambient temperatures are linked to higher baseline body temperatures. Additionally, body temperatures have diurnal and seasonal variation, with low levels at 8 A.M. and during summer and higher levels at 4 P.M. and during winter. Baseline temperatures are also affected by age (lower by 0.02°C for every 10-year increase in age), demographics (African-American women have temperatures 0.052°C higher than white men), and comorbid conditions (cancer is associated with 0.02°C higher temperatures; hypothyroidism is linked to temperatures lower by 0.01°C). After controlling for age, sex, race, vital signs, and comorbidities, an increase in baseline temperature of 0.15°C (or 1 standard deviation) intriguingly translates into a 0.52% absolute increase in 1-year mortality.

Rectal temperatures are generally 0.4°C (0.7°F) higher than oral readings. The lower oral readings are probably attributable to mouth breathing, which is a factor in patients with respiratory infections and rapid breathing. Lower-esophageal temperatures closely reflect core temperature. Tympanic membrane thermometers measure radiant heat from the tympanic membrane and nearby ear canal and display that absolute value (*unadjusted mode*) or a value automatically calculated from the absolute reading on the basis of nomograms relating the radiant temperature measured to actual core temperatures obtained in clinical studies (*adjusted mode*). These measurements, although convenient, may be more variable than directly determined oral or rectal values. Studies in adults show that readings are lower with unadjusted-mode than with adjusted-mode tympanic membrane thermometers and that unadjusted-mode tympanic membrane values are 0.8°C (1.6°F) lower than rectal temperatures.

In women who menstruate, the A.M. temperature is generally lower during the 2 weeks before ovulation; it then rises by ~0.6°C (1°F) with ovulation and stays at that level until menses occur. During the luteal phase, the amplitude of the circadian rhythm remains the same.

FEVER VERSUS HYPERTERMIA

Fever is an elevation of body temperature that exceeds the normal daily variation and occurs *in conjunction with an increase in the hypothalamic set point* (e.g., from 37°C to 39°C). This shift of the set point from “normothermic” to febrile levels very much resembles the resetting of

the home thermostat to a higher level in order to raise the ambient temperature in a room. Once the hypothalamic set point is raised, neurons in the vasomotor center are activated and vasoconstriction commences. The individual first notices vasoconstriction in the hands and feet. Shunting of blood away from the periphery to the internal organs essentially decreases heat loss from the skin, and the person feels cold. For most fevers, body temperature increases by 1–2°C. Shivering, which increases heat production from the muscles, may begin at this time; however, shivering is not required if mechanisms of heat conservation raise blood temperature sufficiently. Nonshivering heat production from the liver also contributes to increasing core temperature. Behavioral adjustments (e.g., putting on more clothing or bedding) help raise body temperature by decreasing heat loss.

The processes of heat conservation (vasoconstriction) and heat production (shivering and increased nonshivering thermogenesis) continue until the temperature of the blood bathing the hypothalamic neurons matches the new “thermostat setting.” Once that point is reached, the hypothalamus maintains the temperature at the febrile level by the same mechanisms of heat balance that function in the afebrile state. When the hypothalamic set point is again reset downward (in response to either a reduction in the concentration of pyrogens or the use of antipyretics), the processes of heat loss through vasodilation and sweating are initiated. Loss of heat by sweating and vasodilation continues until the blood temperature at the hypothalamic level matches the lower setting. Behavioral changes (e.g., removal of clothing) facilitate heat loss.

A fever of >41.5°C (>106.7°F) is called *hyperpyrexia*. This extraordinarily high fever can develop in patients with severe infections but most commonly occurs in patients with central nervous system (CNS) hemorrhages. In the preantibiotic era, fever due to a variety of infectious diseases rarely exceeded 106°F, and there has been speculation that this natural “thermal ceiling” is mediated by neuropeptides functioning as central antipyretics.

In rare cases, the hypothalamic set point is elevated as a result of local trauma, hemorrhage, tumor, or intrinsic hypothalamic malfunction. The term *hypothalamic fever* is sometimes used to describe elevated temperature caused by abnormal hypothalamic function. However, most patients with hypothalamic damage have *subnormal*, not *supranormal*, body temperatures.

Although most patients with elevated body temperature have fever, there are circumstances in which elevated temperature represents not fever but *hyperthermia* (*heat stroke*). Hyperthermia is characterized by an uncontrolled increase in body temperature that exceeds the body's ability to lose heat. The setting of the hypothalamic thermoregulatory center is unchanged. In contrast to fever in infections, hyperthermia does not involve pyrogenic molecules. Exogenous heat exposure and endogenous heat production are two mechanisms by which hyperthermia can result in dangerously high internal temperatures. Excessive heat production can easily cause hyperthermia despite physiologic and behavioral control of body temperature. For example, work or exercise in hot environments can produce heat faster than peripheral mechanisms can lose it. **For a detailed discussion of hyperthermia, see Chap. 465.**

It is important to distinguish between fever and hyperthermia since hyperthermia can be rapidly fatal and characteristically does not respond to antipyretics. In an emergency situation, however, making this distinction can be difficult. For example, in systemic sepsis, fever (hyperpyrexia) can be rapid in onset, and temperatures can exceed 40.5°C (104.9°F). Hyperthermia is often diagnosed on the basis of the events immediately preceding the elevation of core temperature—e.g., heat exposure or treatment with drugs that interfere with thermoregulation. In patients with heat stroke syndromes and in those taking drugs that block sweating, the skin is hot but dry, whereas in fever, the skin can be cold as a consequence of vasoconstriction. Antipyretics do not reduce the elevated temperature in hyperthermia, whereas in fever—and even in hyperpyrexia—adequate doses of either aspirin or acetaminophen usually result in some decrease in body temperature.

PATHOGENESIS OF FEVER

■ PYROGENS

The term *pyrogen* (Greek *pyro*, “fire”) is used to describe any substance that causes fever. *Exogenous* pyrogens are derived from outside the patient; most are microbial products, microbial toxins, or whole microorganisms (including viruses). The classic example of an exogenous pyrogen is the lipopolysaccharide (endotoxin) produced by all gram-negative bacteria. Pyrogenic products of gram-positive organisms include the enterotoxins of *Staphylococcus aureus* and the groups A and B streptococcal toxins, also called *superantigens*. One staphylococcal toxin of clinical importance is that associated with isolates of *S. aureus* from patients with toxic shock syndrome. These products of staphylococci and streptococci cause fever in experimental animals when injected intravenously at concentrations of 1–10 µg/kg. Endotoxin is a highly pyrogenic molecule in humans: when injected intravenously into volunteers, a dose of 2–3 ng/kg produces fever, leukocytosis, acute-phase proteins, and generalized symptoms of malaise.

■ PYROGENIC CYTOKINES

Cytokines are small proteins (molecular mass, 10,000–20,000 Da) that regulate immune, inflammatory, and hematopoietic processes. For example, the elevated leukocytosis seen in several infections with an absolute neutrophilia is attributable to the cytokines interleukin (IL) 1 and IL-6. Some cytokines also cause fever; formerly referred to as *endogenous pyrogens*, they are now called *pyrogenic cytokines*. The pyrogenic cytokines include IL-1, IL-6, tumor necrosis factor (TNF), and ciliary neurotropic factor, a member of the IL-6 family. Fever is a prominent side effect of interferon α therapy. Each pyrogenic cytokine is encoded by a separate gene, and each has been shown to cause fever in laboratory animals and in humans. When injected into humans at low doses (10–100 ng/kg), IL-1 and TNF produce fever; in contrast, for IL-6, a dose of 1–10 µg/kg is required for fever production.

A wide spectrum of bacterial and fungal products induce the synthesis and release of pyrogenic cytokines. However, fever can be a manifestation of disease in the absence of microbial infection. For example, inflammatory processes such as pericarditis, trauma, stroke, and routine immunizations induce the production of IL-1, TNF, and/or IL-6; individually or in combination, these cytokines trigger the hypothalamus to raise the set point to febrile levels.

■ ELEVATION OF THE HYPOTHALAMIC SET POINT BY CYTOKINES

During fever, levels of prostaglandin E₂ (PGE₂) are elevated in hypothalamic tissue and the third cerebral ventricle. The concentrations of PGE₂ are highest near the circumventricular vascular organs (organum vasculosum of lamina terminalis)—networks of enlarged capillaries surrounding the hypothalamic regulatory centers. Destruction of these organs reduces the ability of pyrogens to produce fever. Most studies in animals have failed to show, however, that pyrogenic cytokines pass from the circulation into the brain itself. Thus, it appears that both exogenous pyrogens and pyrogenic cytokines interact with the endothelium of these capillaries and that this interaction is the first step in initiating fever—i.e., in raising the set point to febrile levels.

The key events in the production of fever are illustrated in Fig. 18-1. Myeloid and endothelial cells are the primary cell types that produce pyrogenic cytokines. Pyrogenic cytokines such as IL-1, IL-6, and TNF are released from these cells and enter the systemic circulation. Although these circulating cytokines lead to fever by inducing the synthesis of PGE₂, they also induce PGE₂ in peripheral tissues. The increase in PGE₂ in the periphery accounts for the nonspecific myalgias and arthralgias that often accompany fever. It is thought that some systemic PGE₂ escapes destruction by the lung and gains access to the hypothalamus via the internal carotid. However, it is the elevation of PGE₂ in the brain that starts the process of raising the hypothalamic set point for core temperature.

There are four receptors for PGE₂, and each signals the cell in different ways. Of the four receptors, the third (EP-3) is essential for fever: when the gene for this receptor is deleted in mice, no fever follows the

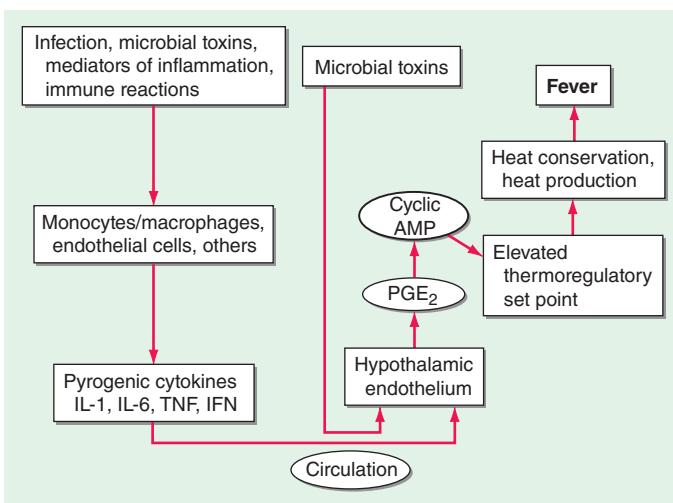


FIGURE 18-1 Chronology of events required for the induction of fever. AMP, adenosine 5'-monophosphate; IFN, interferon; IL, interleukin; PGE₂, prostaglandin E₂; TNF, tumor necrosis factor.

injection of IL-1 or endotoxin. Deletion of the other PGE₂ receptor genes leaves the fever mechanism intact. Although PGE₂ is essential for fever, it is not a neurotransmitter. Rather, the release of PGE₂ from the brain side of the hypothalamic endothelium triggers the PGE₂ receptor on glial cells, and this stimulation results in the rapid release of cyclic adenosine 5'-monophosphate (cAMP), which is a neurotransmitter. As shown in Fig. 18-1, the release of cAMP from glial cells activates neuronal endings from the thermoregulatory center that extend into the area. The elevation of cAMP is thought to account for changes in the hypothalamic set point either directly or indirectly (by inducing the release of neurotransmitters). Distinct receptors for microbial products are located on the hypothalamic endothelium. These receptors are called *Toll-like receptors* and are similar in many ways to IL-1 receptors. IL-1 receptors and Toll-like receptors share the same signal-transducing mechanism. Thus, the direct activation of Toll-like receptors or IL-1 receptors results in PGE₂ production and fever.

■ PRODUCTION OF CYTOKINES IN THE CNS

Cytokines produced in the brain may account for the hyperpyrexia of CNS hemorrhage, trauma, or infection. Viral infections of the CNS induce microglial and possibly neuronal production of IL-1, TNF, and IL-6. In experimental animals, the concentration of a cytokine required to cause fever is several orders of magnitude lower with direct injection into the brain substance or brain ventricles than with systemic injection. Therefore, cytokines produced in the CNS can raise the hypothalamic set point, bypassing the circumventricular organs. CNS cytokines likely account for the hyperpyrexia of CNS hemorrhage, trauma, or infection.

APPROACH TO THE PATIENT

Fever

HISTORY AND PHYSICAL EXAMINATION

There are a range of disease processes that present with fever as a cardinal manifestation, and a thorough history can help distinguish between these broad categories (Table 18-1). The chronology of events preceding fever, including exposure to other symptomatic individuals or to vectors of disease, should be ascertained. Electronic devices for measuring oral, tympanic membrane, or rectal temperatures are reliable, but the same site should be used consistently to monitor a febrile disease. Moreover, physicians should be aware that newborns, elderly patients, patients with chronic hepatic or renal failure, and patients taking glucocorticoids or being treated with an anticytokine may have active disease in the absence of fever because of a blunted febrile response.

TABLE 18-1 Disease Categories That Present with Fever as a Cardinal Sign

Infectious diseases
Autoimmune and noninfectious inflammatory disorders
Cancer
Medication related (e.g., vaccines, drug fever)
Endocrine disorders (e.g., hyperthyroidism)
Intrinsic hypothalamic malfunction

LABORATORY TESTS

The workup should include a complete blood count; a differential count should be performed manually or with an instrument sensitive to the identification of juvenile or band forms, toxic granulations, and Döhle bodies, which are suggestive of bacterial infection. Neutropenia may be present with some viral infections.

Measurement of circulating cytokines in patients with fever is not helpful since levels of cytokines such as IL-1 and TNF in the circulation often are below the detection limit of the assay or do not coincide with fever. However, in patients with low-grade fevers or with suspected occult disease, the most valuable measurements are the C-reactive protein (CRP) level and the erythrocyte sedimentation rate. These markers of inflammatory processes are particularly helpful in detecting occult disease. Measurement of circulating IL-6, which induces CRP, can be useful. However, whereas IL-6 levels may vary during a febrile disease, CRP levels remain elevated. Acute-phase reactants are discussed in [Chap. 304](#).

FEVER IN PATIENTS RECEIVING ANTICYTOKINE THERAPY

Patients receiving long-term treatment with anticytokine-based regimens are at increased risk of infection because of lowered host defenses. For example, latent *Mycobacterium tuberculosis* infection can disseminate in patients receiving anti-TNF therapy. With the increasing use of anticytokines to reduce the activity of IL-1, IL-6, IL-12, IL-17, or TNF in patients with Crohn's disease, rheumatoid arthritis, or psoriasis, the possibility that these therapies blunt the febrile response should be kept in mind.

The blocking of cytokine activity has the distinct clinical drawback of lowering the level of host defenses against both routine bacterial and opportunistic infections such as *M. tuberculosis* and fungal infections. The use of monoclonal antibodies to reduce IL-17 in psoriasis increases the risk of systemic candidiasis.

In nearly all reported cases of infection associated with anticytokine therapy, fever is among the presenting signs. However, the extent to which the febrile response is blunted in these patients remains unknown. Therefore, low-grade fever in patients receiving anticytokine therapies is of considerable concern. The physician should conduct an early and rigorous diagnostic evaluation in these cases. The febrile response is also blunted in patients receiving chronic glucocorticoid therapy or anti-inflammatory agents such as nonsteroidal anti-inflammatory drugs (NSAIDs).

TREATMENT

Fever

THE DECISION TO TREAT FEVER

In deciding whether to treat fever, it is important to remember that fever itself is not an illness: it is an ordinary response to a perturbation of normal host physiology. Most fevers are associated with self-limited infections, such as common viral diseases. The use of antipyretics is not contraindicated in these infections: no significant clinical evidence indicates either that antipyretics delay the resolution of viral or bacterial infections or that fever facilitates recovery from infection or acts as an adjuvant to the immune system. In short, treatment of fever and its symptoms with routine antipyretics

does no harm and does not slow the resolution of common viral and bacterial infections.

However, in bacterial infections, the withholding of antipyretic therapy can be helpful in evaluating the effectiveness of a particular antibiotic, especially in the absence of positive cultures of the infecting organism, and the routine use of antipyretics can mask an inadequately treated bacterial infection. Withholding antipyretics in some cases may facilitate the diagnosis of an unusual febrile disease. Temperature-pulse dissociation (*relative bradycardia*) occurs in typhoid fever, brucellosis, leptospirosis, some drug-induced fevers, and factitious fever. As stated earlier, in newborns, elderly patients, patients with chronic liver or kidney failure, and patients taking glucocorticoids, fever may not be present despite infection. Hypothermia can develop in patients with septic shock.

Some infections have characteristic patterns in which febrile episodes are separated by intervals of normal temperature. For example, *Plasmodium vivax* causes fever every third day, whereas fever occurs every fourth day with *Plasmodium malariae*. Another relapsing fever is related to *Borrelia* infection, with days of fever followed by a several-day afebrile period and then a relapse into additional days of fever. In the Pel-Ebstein pattern, fever lasting 3–10 days is followed by afebrile periods of 3–10 days; this pattern can be classic for Hodgkin's disease and other lymphomas. In cyclic neutropenia, fevers occur every 21 days and accompany the neutropenia. There are also a number of periodic fever syndromes (e.g., familial Mediterranean fever, TNF receptor-associated periodic syndrome [TRAPS]) that differ in their periodicity, duration of attack, constellation of clinical features, genetic causes, and therapies ([Chap. 369](#)). Understanding these clinical differences can help tailor diagnostic testing to confirm the diagnosis and guide therapy.

ANTICYTOKINE THERAPY TO REDUCE FEVER IN AUTOIMMUNE AND AUTOINFLAMMATORY DISEASES

Recurrent fever is documented at some point in most autoimmune diseases and many autoinflammatory diseases, which include the periodic fever syndromes as well as disorders of inflammasomes (e.g., NLRP3, pyrin) and other components of the innate immune system ([Chap. 349](#)). Although fever can be a manifestation of autoimmune diseases, recurrent fevers are characteristic of autoinflammatory diseases, including uncommon diseases such as adult and juvenile Still's disease, familial Mediterranean fever, and hyper-IgD syndrome but also common diseases such as idiopathic pericarditis and gout. In addition to recurrent fevers, neutrophilia and serosal inflammation characterize autoinflammatory diseases. The fevers associated with many of these illnesses are dramatically reduced by blocking of IL-1 activity with anakinra or canakinumab. Anticytokines therefore reduce fever in autoimmune and autoinflammatory diseases. Although fevers in autoinflammatory diseases are mediated by IL-1 β , patients also respond to antipyretics.

MECHANISMS OF ANTIPIRETIC AGENTS

The reduction of fever by lowering of the elevated hypothalamic set point is a direct function of reduction of the PGE₂ level in the thermoregulatory center. The synthesis of PGE₂ depends on the constitutively expressed enzyme cyclooxygenase. The substrate for cyclooxygenase is arachidonic acid released from the cell membrane, and this release is the rate-limiting step in the synthesis of PGE₂. Therefore, inhibitors of cyclooxygenase are potent antipyretics. The antipyretic potency of various drugs is directly correlated with the inhibition of brain cyclooxygenase. Acetaminophen is a poor cyclooxygenase inhibitor in peripheral tissue and lacks noteworthy anti-inflammatory activity; in the brain, however, acetaminophen is oxidized by the P450 cytochrome system, and the oxidized form inhibits cyclooxygenase activity. Moreover, in the brain, the inhibition of another enzyme, COX-3, by acetaminophen may account for the antipyretic effect of this agent. However, COX-3 is not found outside the CNS.

Oral aspirin and acetaminophen are equally effective in reducing fever in humans. NSAIDs such as ibuprofen and specific inhibitors of COX-2 also are excellent antipyretics. Chronic, high-dose

therapy with antipyretics such as aspirin or any NSAID does not reduce normal core body temperature. Thus, PGE₂ appears to play no role in normal thermoregulation.

As effective antipyretics, glucocorticoids act at two levels. First, similar to the cyclooxygenase inhibitors, glucocorticoids reduce PGE₂ synthesis by inhibiting the activity of phospholipase A₂, which is needed to release arachidonic acid from the cell membrane. Second, glucocorticoids block the transcription of the mRNA for the pyrogenic cytokines. Limited experimental evidence indicates that ibuprofen and COX-2 inhibitors reduce IL-1-induced IL-6 production and may contribute to the antipyretic activity of NSAIDs.

REGIMENS FOR THE TREATMENT OF FEVER

The objectives in treating fever are first to reduce the elevated hypothalamic set point and second to facilitate heat loss. Reducing fever with antipyretics also reduces systemic symptoms of headache, myalgias, and arthralgias.

Oral aspirin and NSAIDs effectively reduce fever but can adversely affect platelets and the gastrointestinal tract. Therefore, acetaminophen is preferred as an antipyretic. In children, acetaminophen or oral ibuprofen must be used because aspirin increases the risk of Reye's syndrome. If the patient cannot take oral antipyretics, parenteral preparations of NSAIDs and rectal suppositories of various antipyretics can be used.

Treatment of fever in some patients is highly recommended. Fever increases the demand for oxygen (i.e., for every increase of 1°C over 37°C, there is a 13% increase in oxygen consumption) and can aggravate the condition of patients with preexisting impairment of cardiac, pulmonary, or CNS function. Children with a history of febrile or nonfebrile seizure should be aggressively treated to reduce fever. However, it is unclear what triggers the febrile seizure, and there is no correlation between absolute temperature elevation and onset of a febrile seizure in susceptible children.

In hyperpyrexia, the use of cooling blankets facilitates the reduction of temperature; however, cooling blankets should not be used without oral antipyretics. In hyperpyretic patients with CNS disease or trauma (CNS bleeding), reducing core temperature mitigates the detrimental effects of high temperature on the brain.

For a discussion of treatment for hyperthermia, see Chap. 465.

FURTHER READING

- DINARELLO CA et al: Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. *Nature Rev* 11:633, 2012.
- GATTORNO M et al: Classification criteria for autoinflammatory recurrent fevers. *Ann Rheum Dis* 78:1025, 2019.
- KULLENBERG T et al: Long-term safety profile of anakinra in patients with severe cryopyrin-associated periodic syndromes. *Rheumatology* 55:1499, 2016.
- SAKKAT A et al: Temperature control in critically ill patients with fever: A meta-analysis of randomized controlled trials. *J Crit Care* 61:89, 2021.

19

Fever and Rash

Elaine T. Kaye, Kenneth M. Kaye



The acutely ill patient with fever and rash often presents a diagnostic challenge for physicians, yet the distinctive appearance of an eruption in concert with a clinical syndrome can facilitate a prompt diagnosis and the institution of life-saving therapy or critical infection-control interventions. **Representative images of many of the rashes discussed in this chapter are included in Chap. A1.**

APPROACH TO THE PATIENT

Fever and Rash

A thorough history of patients with fever and rash includes the following relevant information: immune status, medications taken within the previous month, specific travel history, immunization status, exposure to domestic pets and other animals, history of animal (including arthropod) bites, recent dietary exposures, existence of cardiac abnormalities, presence of prosthetic material, recent exposure to ill individuals, and sexual exposures. The history should also include the site of onset of the rash and its direction and rate of spread.

PHYSICAL EXAMINATION

A thorough physical examination entails close attention to the rash, with an assessment and precise definition of its salient features. First, it is critical to determine what *type* of lesions make up the eruption. *Macules* are flat lesions defined by an area of changed color (i.e., a blanchable erythema). *Papules* are raised, solid lesions <5 mm in diameter; *plaques* are lesions >5 mm in diameter with a flat, plateau-like surface; and *nodules* are lesions >5 mm in diameter with a more rounded configuration. *Wheals* (urticaria, hives) are papules or plaques that are pale pink and may appear annular (ringlike) as they enlarge; classic (nonvasculitic) wheals are transient, lasting only 24 h in any defined area. *Vesicles* (<5 mm) and *bullae* (>5 mm) are circumscribed, elevated lesions containing fluid. *Pustules* are raised lesions containing purulent exudate; vesicular processes such as varicella or herpes simplex may evolve to pustules. *Nonpalpable purpura* is a flat lesion that is due to bleeding into the skin. If <3 mm in diameter, the purpuric lesions are termed *petechiae*; if >3 mm, they are termed *ecchymoses*. *Palpable purpura* is a raised lesion that is due to inflammation of the vessel wall (vasculitis) with subsequent hemorrhage. An *ulcer* is a defect in the skin extending at least into the upper layer of the dermis, and an *eschar* (tâche noire) is a necrotic lesion covered with a black crust.

Other pertinent features of rashes include their *configuration* (i.e., annular or target), the *arrangement* of their lesions, and their *distribution* (i.e., central or peripheral).

For further discussion, see Chaps. 56, 58, 122, and 129.

CLASSIFICATION OF RASH

This chapter reviews rashes that reflect systemic disease, but it does not include localized skin eruptions (i.e., cellulitis, impetigo) that may also be associated with fever (Chap. 129). The chapter is not intended to be all-inclusive, but it covers the most important and most common diseases associated with fever and rash. Rashes are classified herein on the basis of lesion morphology and distribution. For practical purposes, this classification system is based on the most typical disease presentations. However, morphology may vary as rashes evolve, and the presentation of diseases with rashes is subject to many variations (Chap. 58). For instance, the classic petechial rash of Rocky Mountain spotted fever (Chap. 187) may initially consist of blanchable erythematous macules distributed peripherally; at times, however, the rash associated with this disease may not be predominantly acral, or no rash may develop at all.

Diseases with fever and rash may be classified by type of eruption: centrally distributed maculopapular, peripheral, confluent desquamative erythematous, vesiculobullous, urticaria-like, nodular, purpuric, ulcerated, or with eschars. Diseases are listed by these categories in Table 19-1, and many are highlighted in the text. However, for a more detailed discussion of each disease associated with a rash, the reader is referred to the chapter dealing with that specific disease. (Reference chapters are cited in the text and listed in Table 19-1.)

CENTRALLY DISTRIBUTED MACULOPAPULAR ERUPTIONS

Centrally distributed rashes, in which lesions are primarily truncal, are the most common type of eruption. The rash of *rubeola* (measles) starts

TABLE 19-1 Diseases Associated with Fever and Rash

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Centrally Distributed Maculopapular Eruptions					
Acute meningococcemia ^a	—	—	—	—	155
Drug reaction with eosinophilia and systemic symptoms (DRESS); also termed drug-induced hypersensitivity syndrome (DIHS) ^b ; Chikungunya ^c ; COVID-19 ^c	—	—	—	—	60
Rubeola (measles, first disease) (Fig. 19-1, Fig. A1-2, Fig. A1-3)	Paramyxovirus	Discrete lesions that become confluent as rash spreads from hairline downward, usually sparing palms and soles; lasts ≥3 days; Koplik's spots	Nonimmune individuals	Cough, conjunctivitis, coryza, severe prostration	205
Rubella (German measles, third disease) (Fig. A1-4)	Togavirus	Spreads from hairline downward, clearing as it spreads; Forchheimer spots	Nonimmune individuals	Adenopathy, arthritis	206
Erythema infectiosum (fifth disease) (Fig. A1-1)	Human parvovirus B19	Bright-red "slapped-cheeks" appearance followed by lacy reticular rash that waxes and wanes over 3 weeks; rarely, papular-purpuric "gloves-and-socks" syndrome on hands and feet	Most common among children 3–12 years old; occurs in winter and spring	Mild fever; arthritis in adults; rash following resolution of fever	197
Exanthem subitum (roseola, sixth disease) (Fig. A1-5)	Human herpesvirus 6 or, less commonly, the closely related human herpesvirus 7	Diffuse maculopapular eruption over trunk and neck; resolves within 2 days	Usually affects children <3 years old	Rash following resolution of fever; similar to Boston exanthem (echovirus 16); febrile seizures may occur	195
Primary HIV infection (Fig. A1-6)	HIV	Nonspecific diffuse macules and papules most commonly on upper thorax, face, collar region; less commonly, urticarial or vesicular lesions; oral or genital ulcers	Individuals recently infected with HIV	Pharyngitis, adenopathy, arthralgias	202
Infectious mononucleosis	Epstein-Barr virus	Diffuse maculopapular eruption (5% of cases; 30–90% if ampicillin is given); urticaria, petechiae in some cases; periorbital edema (50%); palatal petechiae (25%)	Adolescents, young adults	Hepatosplenomegaly, pharyngitis, cervical lymphadenopathy, atypical lymphocytosis, heterophile antibody	194
Other viral exanthems	Echoviruses 2, 4, 9, 11, 16, 19, 25; coxsackieviruses A9, B1, B5; etc.	Wide range of skin findings that may mimic rubella or measles	Affect children more commonly than adults	Nonspecific viral syndromes	204
Exanthematous drug-induced eruption (Fig. A1-7)	Drugs (antibiotics, anticonvulsants, diuretics, etc.)	Intensely pruritic, bright-red macules and papules, symmetric on trunk and extremities; may become confluent	Occurs 2–3 days after exposure in previously sensitized individuals; otherwise, after 2–3 weeks (but can occur anytime, even shortly after drug is discontinued)	Variable findings: fever and eosinophilia	60
Epidemic typhus	<i>Rickettsia prowazekii</i>	Maculopapular eruption appearing in axillae, spreading to trunk and later to extremities; usually spares face, palms, soles; evolves from blanchable macules to confluent eruption with petechiae; rash evanescent in recrudescent typhus (Brill-Zinsser disease)	Exposure to body lice; occurrence of recrudescent typhus as relapse after 30–50 years	Headache, myalgias; mortality rates 10–40% if untreated; milder clinical presentation in recrudescent form	187
Endemic (murine) typhus	<i>Rickettsia typhi</i>	Maculopapular eruption, usually sparing palms, soles	Exposure to rat or cat fleas	Headache, myalgias	187
Scrub typhus	<i>Orientia tsutsugamushi</i>	Diffuse macular rash starting on trunk; eschar at site of mite bite	Endemic in South Pacific, Australia, Asia; transmitted by mites	Headache, myalgias, regional adenopathy; mortality rates up to 30% if untreated	187
Rickettsial spotted fevers (Fig. 19-8)	<i>Rickettsia conorii</i> (boutonneuse fever), <i>Rickettsia australis</i> (North Queensland tick typhus), <i>Rickettsia sibirica</i> (Siberian tick typhus), <i>Rickettsia africae</i> (African tick-bite fever), and others	Eschar common at bite site; maculopapular (rarely, vesicular and petechial) eruption on proximal extremities, spreading to trunk and face	Exposure to ticks; <i>R. conorii</i> in Mediterranean region, India, Africa; <i>R. australis</i> in Australia; <i>R. sibirica</i> in Siberia, Mongolia; <i>R. africae</i> in Africa, Caribbean	Headache, myalgias, regional adenopathy	187

(Continued)

TABLE 19-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Human moncytotropic ehrlichiosis ^a	<i>Ehrlichia chaffeensis</i>	Maculopapular eruption (40% of cases), involves trunk and extremities; may be petechial	Tick-borne; most common in U.S. Southeast, southern Midwest, and mid-Atlantic regions	Headache, myalgias, leukopenia	187
Leptospirosis	<i>Leptospira interrogans</i> and other <i>Leptospira</i> species	Maculopapular eruption; conjunctivitis; scleral hemorrhage in some cases	Exposure to water contaminated with animal urine	Myalgias; aseptic meningitis; fulminant form: icterohemorrhagic fever (Weil's disease)	184
Lyme disease (Fig. A1-8)	<i>Borrelia burgdorferi</i> (sole cause in U.S.), <i>Borrelia afzelii</i> , <i>Borrelia garinii</i>	Papule expanding to erythematous annular lesion with central clearing (erythema migrans; average diameter, 15 cm), sometimes with concentric rings, sometimes with indurated or vesicular center; multiple secondary erythema migrans lesions in some cases	Bite of <i>Ixodes</i> tick vector	Headache, myalgias, chills, photophobia occurring acutely; CNS disease, myocardial disease, arthritis weeks to months later in some cases	186
Southern tick-associated rash illness (STARI, Master's disease)	Unknown (possibly <i>Borrelia lonestari</i> or other <i>Borrelia</i> spirochetes)	Similar to erythema migrans of Lyme disease with several differences, including: multiple secondary lesions less likely; lesions tending to be smaller (average diameter, ~8 cm); central clearing more likely	Bite of tick vector <i>Amblyomma americanum</i> (Lone Star tick); often found in regions where Lyme disease is uncommon, including southern United States	Compared with Lyme disease: fewer constitutional symptoms, tick bite more likely to be recalled; other Lyme disease sequelae lacking	186
Typhoid fever (Fig. A1-9)	<i>Salmonella typhi</i>	Transient, blanchable erythematous macules and papules, 2–4 mm, usually on trunk (rose spots)	Ingestion of contaminated food or water (rare in U.S.)	Variable abdominal pain and diarrhea; headache, myalgias, hepatosplenomegaly	165
Dengue fever ^e (Fig. A1-53)	Dengue virus (4 serotypes; flaviviruses)	Rash in 50% of cases; initially diffuse flushing; midway through illness, onset of maculopapular rash, which begins on trunk and spreads centrifugally to extremities and face; pruritus, hyperesthesia in some cases; after defervescence, petechiae on extremities may occur	Occurs in tropics and subtropics; transmitted by mosquito	Headache; musculoskeletal pain ("breakbone fever"); leukopenia; occasionally biphasic ("saddleback") fever	209
Rat-bite fever (sodoku)	<i>Spirillum minus</i>	Eschar at bite site; then blotchy violaceous or red-brown rash involving trunk and extremities	Rat bite; primarily found in Asia; rare in U.S.	Regional adenopathy; recurrent fevers if untreated	141
Relapsing fever	<i>Borrelia</i> species	Central rash at end of febrile episode; petechiae in some cases	Exposure to ticks or body lice	Recurrent fever, headache, myalgias, hepatosplenomegaly	185
Erythema marginatum (rheumatic fever)	Group A <i>Streptococcus</i>	Erythematous annular papules and plaques occurring as polycyclic lesions in waves over trunk, proximal extremities; evolving and resolving within hours	Patients with rheumatic fever	Pharyngitis preceding polyarthritis, carditis, subcutaneous nodules, chorea	388
Systemic lupus erythematosus (SLE) (Fig. A1-10, Fig. A1-11, Fig. A1-12)	Autoimmune disease	Macular and papular erythema, often in sun-exposed areas; discoid lupus lesions (local atrophy, scale, pigmentary changes); periungual telangiectasis; malar rash; vasculitis sometimes causing urticaria, palpable purpura; oral erosions in some cases	Most common in young to middle-aged women; flares precipitated by sun exposure	Arthritis; cardiac, pulmonary, renal, hematologic, and vasculitic disease	359
Still's disease (Fig. A1-13)	Autoimmune disease	Transient 2- to 5-mm erythematous papules appearing at height of fever on trunk, proximal extremities; lesions evanescent	Children and young adults	High spiking fever, polyarthritis, splenomegaly; erythrocyte sedimentation rate >100 mm/h	—
African trypanosomiasis (Fig. A1-47)	<i>Trypanosoma brucei rhodesiense/gambiense</i>	Blotchy or annular erythematous macular and papular rash (trypanid), primarily on trunk; pruritus; chancre at site of tsetse fly bite may precede rash by several weeks	Tsetse fly bite in eastern (<i>T. brucei rhodesiense</i>) or western (<i>T. brucei gambiense</i>) Africa	Hemolymphatic disease followed by meningoencephalitis; Winterbottom's sign (posterior cervical lymphadenopathy) (<i>T. brucei gambiense</i>)	227
Arcanobacterial pharyngitis	<i>Arcanobacterium (Corynebacterium) haemolyticum</i>	Diffuse, erythematous, maculopapular eruption involving trunk and proximal extremities; may desquamate	Children and young adults	Exudative pharyngitis, lymphadenopathy	150

(Continued)

TABLE 19-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
West Nile virus infection	West Nile virus	Maculopapular eruption involving the trunk, extremities, and head or neck; rash in 20–50% of cases	Mosquito bite; rarely, blood transfusion or transplanted organ	Headache, weakness, malaise, myalgia, neuroinvasive disease (encephalitis, meningitis, flaccid paralysis)	209
Zika virus infection (Fig. A1-51)	Zika virus	Pruritic macular and papular erythema; rash may begin on trunk and descend to lower body; conjunctival injection; palatal petechiae may occur	Mosquito bite; sexual transmission or blood transfusion less common	Arthralgia (especially of small joints), myalgia, lymphadenopathy, headache, low-grade fever; illness in pregnancy may cause severe birth defects, including microcephaly; neurologic complications, including Guillain-Barré, may occur	209
Peripheral Eruptions					
Chronic meningococcemia, disseminated gonococcal infection, ^a human parvovirus B19 infection, ^f MIRM ^g	—	—	—	—	155, 156, 197
Rocky Mountain spotted fever (Fig. 19-2, Fig. A1-16)	<i>Rickettsia rickettsii</i>	Rash beginning on wrists and ankles and spreading centripetally; appears on palms and soles later in disease; lesion evolution from blanchable macules to petechiae	Tick vector; widespread but more common in southeastern and southwest-central U.S.	Headache, myalgias, abdominal pain; mortality rates up to 40% if untreated	187
Secondary syphilis (Figs. A1-18, Fig. A1-19, Fig. A1-20, Fig. A1-21)	<i>Treponema pallidum</i>	Coincident primary chancre in 10% of cases; copper-colored, scaly papular eruption, diffuse but prominent on palms and soles; rash never vesicular in adults; condyloma latum, mucous patches, and alopecia in some cases	Sexually transmitted	Fever, constitutional symptoms	182
Chikungunya fever (Fig. A1-54)	Chikungunya virus	Maculopapular eruption; typically occurs on trunk, but also occurs on extremities and face	<i>Aedes aegypti</i> and <i>A. albopictus</i> mosquito bites; tropical and subtropical regions	Severe polyarticular, migratory arthralgias, especially involving small joints (e.g., hands, wrists, ankles)	209
Hand-foot-and-mouth disease (Fig. A1-22)	Coxsackievirus A16 and enterovirus 71 most common causes; coxsackievirus A6 associated with atypical syndrome	Tender vesicles, erosions in mouth; 0.25-cm papules on hands and feet with rim of erythema evolving into tender vesicles; shedding of nails (onychomadesis) can occur 1–2 months after acute illness; coxsackievirus A6 lesions may also be maculopapular, petechial, purpuric, or erosive; atypical form often extends to perioral area, extremities, trunk, buttocks, genitals, and areas affected by eczema (eczema coxsackium)	Summer and fall; primarily children <10 years old; multiple family members; coxsackievirus A6 infection also occurs in young adults	Transient fever; enterovirus 71 can be associated with brain stem encephalitis, flaccid paralysis resembling polio, or aseptic meningitis	204
Erythema multiforme (EM) (Fig. A1-24)	Infection, drugs, idiopathic causes	Target lesions (central erythema surrounded by area of clearing and another rim of erythema) up to 2 cm; symmetric on knees, elbows, palms, soles; spreads centripetally; papular, sometimes vesicular; when extensive and involving mucous membranes, termed EM major	Herpes simplex virus or <i>Mycoplasma pneumoniae</i> infection; drug intake (i.e., sulfa, phenytoin, penicillin)	50% of patients <20 years old; fever more common in most severe form, EM major, which can be confused with Stevens-Johnson syndrome (but EM major lacks prominent skin sloughing)	— ^b
Rat-bite fever (Haverhill fever)	<i>Streptobacillus moniliformis</i>	Maculopapular eruption over palms, soles, and extremities; tends to be more severe at joints; eruption sometimes becoming generalized; may be purpuric; may desquamate	Rat bite, ingestion of contaminated food	Myalgias; arthritis (50%); fever recurrence in some cases	141

(Continued)

TABLE 19-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Bacterial endocarditis (Fig. A1-23)	<i>Streptococcus</i> , <i>Staphylococcus</i> , etc.	<i>Subacute course</i> (e.g., viridans streptococci): Osler's nodes (tender pink nodules on finger or toe pads); petechiae on skin and mucosa; splinter hemorrhages. <i>Acute course</i> (e.g., <i>Staphylococcus aureus</i>): Janeway lesions (painless erythematous or hemorrhagic macules, usually on palms and soles)	Abnormal heart valve (e.g., viridans streptococci), intravenous drug use	New or changing heart murmur	128
COVID-19 (Fig. A1-57)	SARS-CoV-2	<i>Mild or asymptomatic COVID-19</i> : Pernio (macules, papules, or plaques that are tender, erythematous/violaceous; acral, feet more common than hands); <i>Moderate/severe COVID-19</i> : vesicles, urticaria, maculopapular erythema; often pruritic; occur on trunk, extremities; <i>Severe COVID-19</i> : Retiform purpura (net-like, purple patches/plaques often with necrosis); lesions often asymptomatic; occur on extremities, buttocks; <i>Multisystem inflammatory syndrome in children (MIS-C)</i> : findings similar to Kawasaki disease	Infection with SARS-CoV-2; MIS-C in older children/adolescents	Ranging from asymptomatic to mild/moderate with loss of taste/smell, pharyngitis, cough, fever, to severe with dyspnea, ARDS; complications include thrombosis, especially with retiform purpura; lesions may be delayed compared to other COVID-19 symptoms; MIS-C occurs ~2-6 weeks following acute (often asymptomatic) infection	

Confluent Desquamative Erythemas

Scarlet fever (second disease) (Fig. A1-25)	Group A <i>Streptococcus</i> (pyrogenic exotoxins A, B, C)	Diffuse blanchable erythema beginning on face and spreading to trunk and extremities; circumoral pallor; "sandpaper" texture to skin; accentuation of linear erythema in skin folds (Pastia's lines); enanthem of white evolving into red "strawberry" tongue; desquamation in second week	Most common among children 2–10 years old; usually follows group A streptococcal pharyngitis	Fever, pharyngitis, headache	148
Kawasaki disease (Fig. A1-29)	Idiopathic	Rash similar to scarlet fever (scarlatiniform) or EM; fissuring of lips, strawberry tongue; conjunctivitis; edema of hands, feet; desquamation later in disease	Children <8 years old	Cervical adenopathy, pharyngitis, coronary artery vasculitis	58, 363
Streptococcal toxic shock syndrome	Group A <i>Streptococcus</i> (associated with pyrogenic exotoxin A and/or B or certain M types)	When present, rash often scarlatiniform	May occur in setting of severe group A streptococcal infections (e.g., necrotizing fasciitis, bacteremia, pneumonia)	Multiorgan failure, hypotension; mortality rate 30%	148
Staphylococcal toxic shock syndrome	<i>S. aureus</i> (toxic shock syndrome toxin 1, enterotoxins B and others)	Diffuse erythema involving palms; pronounced erythema of mucosal surfaces; conjunctivitis; desquamation 7–10 days into illness	Colonization with toxin-producing <i>S. aureus</i>	Fever >39°C (>102°F), hypotension, multiorgan dysfunction	147
Staphylococcal scalded-skin syndrome (Fig. 19-3, Fig. A1-28)	<i>S. aureus</i> , phage group II	Diffuse tender erythema, often with bullae and desquamation; Nikolsky's sign	Colonization with toxin-producing <i>S. aureus</i> ; occurs in children <10 years old (termed <i>Ritter's disease</i> in neonates) or adults with renal dysfunction	Irritability; nasal or conjunctival secretions	147
Exfoliative erythroderma syndrome (Fig. A1-27)	Underlying psoriasis, eczema, drug eruption, mycosis fungoides	Diffuse erythema (often scaling) interspersed with lesions of underlying condition	Usually occurs in adults over age 50; more common among men	Fever, chills (i.e., difficulty with thermoregulation); lymphadenopathy	58, 60
DRESS (drug-induced hypersensitivity syndrome [DIHS]) (Fig. A1-48)	Aromatic anticonvulsants; other drugs, including sulfonamides, minocycline	Maculopapular eruption (mimicking exanthematous drug rash), sometimes progressing to exfoliative erythroderma; profound edema, especially facial; pustules may occur	Individuals genetically unable to detoxify arene oxides (anticonvulsant metabolites), patients with slow N-acetylating capacity (sulfonamides)	Lymphadenopathy, multiorgan failure (especially hepatic), eosinophilia, atypical lymphocytes; mimics sepsis	60
Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) (Fig. A1-26)	Drugs (80% of cases; often allopurinol, anticonvulsants, antibiotics), infection, idiopathic factors	Erythematous and purpuric macules, sometimes targetoid, or diffuse erythema progressing to bullae, with sloughing and necrosis of entire epidermis; Nikolsky's sign; involves mucosal surfaces; TEN (>30% epidermal necrosis) is maximal form; SJS involves <10% of epidermis; SJS/TEN overlap involves 10–30% of epidermis	Uncommon among children; more common among patients with HIV infection, systemic lupus erythematosus, certain HLA types, or slow acetylators	Dehydration, sepsis sometimes resulting from lack of normal skin integrity; mortality rates up to 30%	60

(Continued)

TABLE 19-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Vesiculobullous or Pustular Eruptions					
Hand-foot-and-mouth syndrome ^a ; staphylococcal scalded-skin syndrome ^b ; TEN ^b ; DRESS ^b ; COVID-19 ^c	—	—	—	—	— ^d
Varicella (chickenpox) (Fig. 19-4, Fig. A1-30)	Varicella-zoster virus (VZV)	Macules (2–3 mm) evolving into papules, then vesicles (sometimes umbilicated), on an erythematous base ("dewdrops on a rose petal"); pustules then forming and crusting; lesions appearing in crops; may involve scalp, mouth; intensely pruritic	Usually affects children; 10% of adults susceptible; most common in late winter and spring; incidence down by 90% in U.S. as a result of varicella vaccination	Malaise; generally mild disease in healthy children; more severe disease with complications in adults and immunocompromised children	193
Pseudomonas "hot-tub" folliculitis (Fig. A1-55)	<i>Pseudomonas aeruginosa</i>	Pruritic erythematous follicular, papular, vesicular, or pustular lesions that may involve axillae, buttocks, abdomen, and especially areas occluded by bathing suits; can manifest as tender isolated nodules on palmar or plantar surfaces (the latter designated " <i>Pseudomonas</i> hot-foot syndrome")	Bathers in hot tubs or swimming pools; occurs in outbreaks	Earache, sore eyes and/or throat; fever may be absent; generally self-limited	164
Variola (smallpox) (Fig. A1-50)	Variola major virus	Red macules on tongue and palate evolving to papules and vesicles; skin macules evolving to papules, then vesicles, then pustules over 1 week, with subsequent lesion crusting; lesions initially appearing on face and spreading centrifugally from trunk to extremities; differs from varicella in that (1) skin lesions in any given area are at same stage of development and (2) there is a prominent distribution of lesions on face and extremities (including palms, soles)	Nonimmune individuals exposed to smallpox	Prodrome of fever, headache, backache, myalgias; vomiting in 50% of cases	S3
Primary herpes simplex virus (HSV) infection	HSV	Erythema rapidly followed by hallmark painful <i>grouped vesicles</i> that may evolve into pustules that ulcerate, especially on mucosal surfaces; lesions at site of inoculation: commonly gingivostomatitis for HSV-1 and genital lesions for HSV-2; recurrent disease milder (e.g., herpes labialis does not involve oral mucosa)	Primary infection most common among children and young adults for HSV-1 and among sexually active young adults for HSV-2; no fever in recurrent infection	Regional lymphadenopathy	192
Disseminated herpesvirus infection (Fig. A1-31)	VZV or HSV	Generalized vesicles that can evolve to pustules and ulcerations; individual lesions similar for VZV and HSV. <i>Zoster cutaneous dissemination</i> : >25 lesions extending outside involved dermatome. HSV: extensive, progressive mucocutaneous lesions that may occur in absence of dissemination, sometimes disseminate in eczematous skin (<i>eczema herpeticum</i>); HSV visceral dissemination may occur with only localized mucocutaneous disease; in disseminated neonatal disease, skin lesions diagnostically helpful when present, but rash absent in a substantial minority of cases	Patients with immunosuppression, eczema; neonates	Visceral organ involvement (e.g., liver, lungs) in some cases; neonatal disease particularly severe	138, 192, 193
Rickettsialpox (Fig. A1-33)	<i>Rickettsia akari</i>	Eschar found at site of mite bite; generalized rash involving face, trunk, extremities; may involve palms and soles; <100 papules and plaques (2–10 mm); centers of papules develop vesicles or pustules	Seen in urban settings; transmitted by mouse mites	Headache, myalgias, regional adenopathy; mild disease	187
Acute generalized exanthematous pustulosis (Fig. A1-49)	Drugs (mostly anticonvulsants or antimicrobials); also viral	Tiny, sterile, nonfollicular pustules on erythematous, edematous skin; begins on face and in body folds, then becomes generalized	Appears 2–21 days after start of drug therapy, depending on whether patient has been sensitized	Acute fever, pruritus, leukocytosis	60

(Continued)

TABLE 19-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Disseminated <i>Vibrio vulnificus</i> infection	<i>V. vulnificus</i>	Erythematous lesions evolving into hemorrhagic bullae and then into necrotic ulcers	Patients with cirrhosis, diabetes, renal failure; exposure by ingestion of contaminated saltwater, seafood	Hypotension; mortality rate 50%	168
Ecthyma gangrenosum (Fig. A1-34)	<i>P. aeruginosa</i> , other gram-negative rods, fungi	Indurated plaque evolving into hemorrhagic bulla or pustule that sloughs, resulting in eschar formation; erythematous halo; most common in axillary, groin, perianal regions	Usually affects neutropenic patients; occurs in up to 28% of individuals with <i>Pseudomonas</i> bacteremia	Clinical signs of sepsis	164
Mycoplasma-induced rash and mucositis (MIRM)	<i>Mycoplasma pneumoniae</i>	Severe mucositis of at least two sites (e.g., oropharynx, ocular, genital) with nearly universal hemorrhagic crusting of lips; sparse, vesiculobullous, or atypical targetoid rash over <10% of body; lesions typically on extremities but can be truncal; rash sometimes absent (MIRM sine rash)	More common in males; usually children (mean age 11–12 years old)	Evidence of <i>M. pneumoniae</i> infection (typically pneumonia); good prognosis; distinct from SJS/TEN; rarely <i>Chlamydophila pneumoniae</i> can cause similar syndrome	
Urticaria-Like Eruptions					
COVID-19 ^c					
Urticular vasculitis (Fig. 19-5, Fig. A1-35)	Serum sickness, often due to infection (including acute hepatitis B, enteroviral, parasitic), drugs; connective tissue disease	Erythematous, edematous “urticaria-like” plaques, pruritic or burning; unlike urticaria: typical lesion duration >24 h (up to 5 days) and lack of complete lesion blanching with compression due to hemorrhage	Patients with serum sickness (including acute hepatitis B), connective tissue disease	Fever variable; arthralgias/arthritis	363 ^d
Nodular Eruptions					
Disseminated infection (Fig. 19-6, Fig. A1-36, Fig. A1-37, Fig. A1-38)	Fungal infections (e.g., candidiasis, histoplasmosis, cryptococcosis, sporotrichosis, coccidioidomycosis); mycobacteria	Subcutaneous nodules (up to 3 cm); fluctuance, draining common with mycobacteria; necrotic nodules (extremities, periorbital or nasal regions) common with <i>Aspergillus</i> , <i>Mucor</i>	Immunocompromised hosts (e.g., bone marrow transplant recipients, patients undergoing chemotherapy, HIV-infected patients)	Features vary with organism	— ^h
Erythema nodosum (septal panniculitis) (Fig. A1-39)	Infections (e.g., streptococcal, fungal, mycobacterial, yersinial); drugs (e.g., sulfas, penicillins, oral contraceptives); sarcoidosis; idiopathic causes	Large, violaceous, nonulcerative, subcutaneous nodules; exquisitely tender; usually on lower legs but also on upper extremities	More common among females 15–30 years old	Arthralgias (50%); features vary with associated condition	— ^h
Sweet syndrome (acute febrile neutrophilic dermatosis) (Fig. A1-40)	<i>Yersinia</i> infection; upper respiratory infection; inflammatory bowel disease; pregnancy; malignancy (usually hematologic); drugs (G-CSF)	Tender red or blue edematous nodules giving impression of vesiculation; usually on face, neck, upper extremities; when on lower extremities, may mimic erythema nodosum	More common among women and among persons 30–60 years old; 20% of cases associated with malignancy (men and women equally affected in this group)	Headache, arthralgias, leukocytosis	58
Bacillary angiomatosis	<i>Bartonella henselae</i> , <i>B. quintana</i>	Many forms, including erythematous, smooth vascular nodules; friable, exophytic lesions; erythematous plaques (may be dry, scaly); subcutaneous nodules (may be erythematous)	Immunosuppressed individuals, especially those with advanced HIV infection	Peliosis of liver and spleen in some cases; lesions sometimes involving multiple organs; bacteremia	172
Purpuric Eruptions					
Rocky Mountain spotted fever, rat-bite fever, endocarditis ^e ; epidemic typhus ^f ; dengue fever ^{e,f} ; human parvovirus B19 infection ^g ; COVID-19 ^c	—	—	—	—	— ^h
Acute meningococcemia	<i>Neisseria meningitidis</i>	Initially pink maculopapular lesions evolving into petechiae; petechiae rapidly becoming numerous, sometimes enlarging and becoming vesicular; trunk, extremities most commonly involved; may appear on face, hands, feet; may include purpura fulminans (see below) reflecting DIC	Most common among children, individuals with asplenia or terminal complement component deficiency (C5–C8)	Hypotension, meningitis (sometimes preceded by upper respiratory infection)	155

(Continued)

TABLE 19-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Purpura fulminans (Fig. 19-7, Fig. A1-41)	Severe DIC	Large ecchymoses with sharply irregular shapes evolving into hemorrhagic bullae and then into black necrotic lesions	Individuals with sepsis (e.g., involving <i>N. meningitidis</i>), malignancy, or massive trauma; asplenic patients at high risk for sepsis	Hypotension	155, 304
Chronic meningococcemia (Fig. A1-42)	<i>N. meningitidis</i>	Variety of recurrent eruptions, including pink maculopapular; nodular (usually on lower extremities); petechial (sometimes developing vesicular centers); purpuric areas with pale blue-gray centers	Individuals with complement deficiencies	Fevers, sometimes intermittent; arthritis, myalgias, headache	155
Disseminated gonococcal infection (Fig. A1-43)	<i>Neisseria gonorrhoeae</i>	Papules (1–5 mm) evolving over 1–2 days into hemorrhagic pustules with gray necrotic centers; hemorrhagic bullae occurring rarely; lesions (usually <40) distributed peripherally near joints (more commonly on upper extremities)	Sexually active individuals (more often females), some with complement deficiency	Low-grade fever, tenosynovitis, arthritis	156
Enteroviral petechial rash	Usually echovirus 9 or coxsackievirus A9	Disseminated petechial lesions (may also be maculopapular, vesicular, or urticarial)	Often occurs in outbreaks	Pharyngitis, headache; aseptic meningitis with echovirus 9	204
Viral hemorrhagic fever	Arenaviruses, bunyaviruses, filoviruses (including Ebola), flaviviruses (including dengue)	Petechial rash	Residence in or travel to endemic areas, other virus exposure	Triad of fever, shock, hemorrhage from mucosa or gastrointestinal tract	209, 210
Thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome	Idiopathic, bloody diarrhea caused by Shiga toxin-generating bacteria (e.g., <i>Escherichia coli</i> O157:H7), deficiency in ADAMTS13 (cleaves von Willebrand factor), drugs (e.g., quinine, chemotherapy, immunosuppression)	Petechiae	Individuals with <i>E. coli</i> O157:H7 gastroenteritis (especially children), cancer chemotherapy, HIV infection, autoimmune diseases, pregnant/postpartum women, those with ADAMTS13 deficiency	Fever (not always present), microangiopathic hemolytic anemia, thrombocytopenia, renal dysfunction, neurologic dysfunction; coagulation studies normal	58, 100, 115, 161, 166
Cutaneous small-vessel vasculitis (leukocytoclastic vasculitis) (Fig. A1-44)	Infections (including group A streptococcal infection, hepatitis B or C), drugs, idiopathic factors	Palpable purpuric lesions appearing in crops on legs or other dependent areas; may become vesicular or ulcerative	Occurs in a wide spectrum of diseases, including connective tissue disease, cryoglobulinemia, malignancy, Henoch-Schönlein purpura (HSP); more common among children	Fever (not always present), malaise, arthralgias, myalgias; systemic vasculitis in some cases; renal, joint, and gastrointestinal involvement common in HSP	58
Eruptions with Ulcers and/or Eschars					
Scrub typhus, rickettsial spotted fevers, rat-bite fever, African trypanosomiasis ^a , rickettsialpox, ecthyma gangrenosum ^b	—	—	—	—	— ^c
Tularemia (Fig. A1-45, Fig. A1-46)	<i>Francisella tularensis</i>	Ulceroglandular form: erythematous, tender papule evolves into necrotic, tender ulcer with raised borders; in 35% of cases, eruptions (maculopapular, vesiculopapular, acneiform, or urticarial; erythema nodosum; or EM) may occur	Exposure to ticks, biting flies, infected animals	Fever, headache, lymphadenopathy	170
Anthrax (Fig. A1-52)	<i>Bacillus anthracis</i>	Pruritic papule enlarging and evolving into a 1- by 3-cm painless ulcer surrounded by vesicles and then developing a central eschar with edema; residual scar	Exposure to infected animals or animal products, other exposure to anthrax spores	Lymphadenopathy, headache	S3

^aSee "Purpuric Eruptions." ^bSee "Confluent Desquamative Erythemas." ^cSee "Peripheral Eruptions." ^dRash is rare in human granulocytotropic ehrlichiosis or anaplasmosis (caused by *Anaplasma phagocytophylum*; most common in the upper midwestern and northeastern United States). ^eSee "Viral hemorrhagic fever" under "Purpuric Eruptions" for dengue hemorrhagic fever/dengue shock syndrome. ^fSee "Centrally Distributed Maculopapular Eruptions." ^gSee "Vesiculobullous or Pustular Eruptions." ^hSee etiology-specific chapters.

Abbreviations: CNS, central nervous system; DIC, disseminated intravascular coagulation; G-CSF, granulocyte colony-stimulating factor; HLA, human leukocyte antigen.



FIGURE 19-1 Centrally distributed, maculopapular eruption on the trunk in a patient with measles. (From EJ Mayeaux Jr et al: Measles, in Usatine RP et al [eds]: *Color Atlas and Synopsis of Family Medicine*, 3rd ed. New York, McGraw-Hill, 2019, p. 797, Figure 132-2. Reproduced with permission from Richard P. Usatine, MD.)

at the hairline 2–3 days into the illness and moves down the body, typically sparing the palms and soles (Fig. 19-1; see also Fig. A1-3) (Chap. 205). It begins as discrete erythematous lesions, which become confluent as the rash spreads. Koplik's spots (1- to 2-mm white or bluish lesions with an erythematous halo on the buccal mucosa) (Fig. A1-2) are pathognomonic for measles and are generally seen during the first 2 days of symptoms. They should not be confused with Fordyce's spots (ectopic sebaceous glands), which have no erythematous halos and are found in the mouth of healthy individuals. Koplik's spots may briefly overlap with the measles exanthem.

Rubella (German measles) (Fig. A1-4) also spreads from the hairline downward; unlike that of measles, however, the rash of rubella tends to clear from originally affected areas as it migrates, and it may be pruritic (Chap. 206). Forchheimer spots (palatal petechiae) may develop but are nonspecific because they also develop in *infectious mononucleosis* (Chap. 194), *scarlet fever* (Chap. 148), and *Zika virus infection* (Chap. 209) (Fig. A1-51D). Postauricular and suboccipital adenopathy and arthritis are common among adults with rubella. Exposure of pregnant women to ill individuals should be avoided, as rubella causes severe congenital abnormalities. Numerous strains of *enteroviruses* (Chap. 204), primarily echoviruses and coxsackieviruses, cause non-specific syndromes of fever and eruptions that may mimic rubella or measles. Patients with *infectious mononucleosis* caused by Epstein-Barr virus (Chap. 194) or with *primary HIV infection* (Fig. A1-6; see also Chapter 202) may exhibit pharyngitis, lymphadenopathy, and a non-specific maculopapular exanthem.

The rash of *erythema infectiosum* (fifth disease), which is caused by human parvovirus B19, primarily affects children 3–12 years old; it develops after fever has resolved as a bright blanchable erythema on the cheeks ("slapped cheeks") (Fig. A1-1A) with perioral pallor (Chap. 197). A more diffuse rash (often pruritic) appears the next day on the trunk and extremities and then rapidly develops into a lacy reticular eruption (Fig. A1-1B) that may wax and wane (especially with temperature change) over 3 weeks. Adults with fifth disease often have arthritis, and fetal hydrops can develop in association with this condition in pregnant women.

Exanthem subitum (roseola) is caused by human herpesvirus 6, or less commonly by the closely related human herpesvirus 7, and is most

common among children <3 years of age (Chap. 195). As in erythema infectiosum, the rash usually appears after fever has subsided. It consists of 2- to 3-mm rose-pink macules and papules that coalesce only rarely, occur initially on the trunk (Fig. A1-5) and sometimes on the extremities (sparing the face), and fade within 2 days.

Although drug reactions have many manifestations, including urticaria, exanthematic drug-induced eruptions (Chap. 60) (Fig. A1-7) are most common and are often difficult to distinguish from viral exanthems. Eruptions elicited by drugs are usually more intensely erythematous and pruritic than viral exanthems, but this distinction is not reliable. A history of new medications and an absence of prostration may help to distinguish a drug-related rash from an eruption of another etiology. Rashes may persist for up to 2 weeks after administration of the offending agent is discontinued. Certain populations are more prone than others to drug rashes. Of HIV-infected patients, 50–60% develop a rash in response to sulfa drugs; 30–90% of patients with mononucleosis due to Epstein-Barr virus develop a rash when given ampicillin.

Rickettsial illnesses (Chap. 187) should be considered in the evaluation of individuals with centrally distributed maculopapular eruptions. The usual setting for *epidemic typhus* is a site of war or natural disaster in which people are exposed to body lice. Endemic typhus or *leptospirosis* (the latter caused by a spirochete) (Chap. 184) may be seen in urban environments where rodents proliferate. Outside the United States, other rickettsial diseases cause a spotted-fever syndrome and should be considered in residents of or travelers to endemic areas. Similarly, *typhoid fever*, a nonrickettsial disease caused by *Salmonella typhi* (Chap. 165) (Fig. A1-9), is usually acquired during travel outside the United States. *Dengue fever* (Fig. A1-53), caused by a mosquito-transmitted flavivirus, occurs in tropical and subtropical regions of the world (Chap. 209).

Some centrally distributed maculopapular eruptions have distinctive features. Erythema migrans (Fig. A1-8), the rash of *Lyme disease* (Chap. 186), typically manifests as single or multiple annular lesions. Untreated erythema migrans lesions usually fade within a month but may persist for more than a year. *Southern tick-associated rash illness* (STARI) (Chap. 186) has an erythema migrans-like rash, but is less severe than Lyme disease and often occurs in regions where Lyme is not endemic. Erythema marginatum, the rash of *acute rheumatic fever* (Chap. 359), has a distinctive pattern of enlarging and shifting transient annular lesions.

Collagen vascular diseases may cause fever and rash. Patients with *systemic lupus erythematosus* (Chap. 356) typically develop a sharply defined, erythematous eruption in a butterfly distribution on the cheeks (malar rash) (Fig. A1-10) as well as many other skin manifestations (Figs. A1-11, A1-12). *Still's disease* presents as an evanescent, salmon-colored rash on the trunk and proximal extremities that coincides with fever spikes (Fig. A1-13).

Hemophagocytic lymphohistiocytosis may be familial or triggered by infection, autoimmunity, or neoplasia. Cutaneous manifestations are protean and can present as an erythematous maculopapular eruption, pyoderma gangrenosum, purpura, panniculitis, or Stevens Johnson syndrome.

Zika virus is a mosquito-transmitted flavivirus that is associated with severe birth defects (Chap. 209). Zika is widespread among tropical and subtropical regions of the world. The eruption of Zika virus infection (Fig. A1-51A, A1-51B) is typically pruritic and often accompanied by conjunctival injection (Fig. A1-51C).

■ PERIPHERAL ERUPTIONS

These rashes are alike in that they are most prominent peripherally or begin in peripheral (acral) areas before spreading centripetally. Early diagnosis and therapy are critical in *Rocky Mountain spotted fever* (Chap. 187) because of its grave prognosis if untreated. Lesions (Fig. 19-2; see also Fig. A1-16) evolve from macular to petechial, start on the wrists and ankles, spread centripetally, and appear on the palms and soles only later in the disease. The rash of *secondary syphilis* (Chap. 182), which may be generalized (Fig. A1-18) but is prominent on the palms and soles (Fig. A1-19), should be considered in the differential diagnosis of pityriasis rosea, especially in sexually active patients. *Chikungunya fever* (Chap. 209), which is transmitted by mosquito bite



FIGURE 19-2 Peripheral eruption on the wrist and palm exhibiting erythematous macules in the process of evolving into petechial lesions in a patient with Rocky Mountain spotted fever. (From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 562, Figure 25-50; with permission.)

in tropical and subtropical regions, is associated with a maculopapular eruption (Fig. A1-54) and severe polyarticular small-joint arthralgias. *Hand-foot-and-mouth disease* (Chap. 204), most commonly caused by coxsackievirus A16 or enterovirus 71, is distinguished by tender vesicles distributed on the hands and feet and in the mouth (Fig. A1-22); coxsackievirus A6 causes an atypical syndrome with more extensive lesions. The classic target lesions of *erythema multiforme* (Fig. A1-24) appear symmetrically on the elbows, knees, palms, soles, and face. In severe cases, these lesions spread diffusely and involve mucosal surfaces. Lesions may develop on the hands and feet in *endocarditis* (Fig. A1-23) (Chap. 128). Pernio, tender violaceous lesions that are acral (Fig. A1-57), occur most commonly on the feet, in asymptomatic or mild COVID-19. Vesicles, urticaria, or maculopapular eruptions, often pruritic, may occur on the trunk and extremities in moderate or severe disease, while retiform purpura occurs on the extremities and buttocks in severe COVID-19.

■ CONFLUENT DESQUAMATIVE ERYTHEMAS

These eruptions consist of diffuse erythema frequently followed by desquamation. The eruptions caused by group A *Streptococcus* or *Staphylococcus aureus* are toxin-mediated. *Scarlet fever* (Chap. 148) (Fig. A1-25) usually follows pharyngitis; patients have a facial flush, a “strawberry” tongue, and accentuated petechiae in body folds (Pastia’s lines). *Kawasaki disease* (Fig. A1-29) (Chaps. 58 and 363) presents in the pediatric population as fissuring of the lips, a strawberry tongue, conjunctivitis, adenopathy, and sometimes cardiac abnormalities. *Streptococcal toxic shock syndrome* (Chap. 148) manifests with hypotension, multiorgan failure, and, often, a severe group A streptococcal infection (e.g., necrotizing fasciitis). *Staphylococcal toxic shock syndrome* (Chap. 147) also presents with hypotension and multiorgan failure, but usually only *S. aureus* colonization—not a severe *S. aureus* infection—is documented. *Staphylococcal scalded-skin syndrome* (Fig. A1-28) (Chap. 147) is seen primarily in children and in immunocompromised adults. Generalized erythema is often evident during

the prodrome of fever and malaise; profound tenderness of the skin is distinctive. In the exfoliative stage, the skin can be induced to form bullae with light lateral pressure (Nikolsky’s sign) (Fig. 19-3). In a mild form, a scarlatiniform eruption mimics scarlet fever, but the patient does not exhibit a strawberry tongue or circumoral pallor. In contrast to the staphylococcal scalded-skin syndrome, in which the cleavage plane is superficial in the epidermis, *toxic epidermal necrolysis* (Chap. 60), a maximal variant of *Stevens-Johnson syndrome*, involves sloughing of the entire epidermis (Fig. A1-26), resulting in severe disease. *Exfoliative erythroderma syndrome* (Chaps. 58 and 60) is a serious reaction associated with systemic toxicity that is often due to eczema, psoriasis (Fig. A1-27), a drug reaction, or mycosis fungoides. *Drug rash with eosinophilia and systemic symptoms* (DRESS), often due to antiepileptic and antibiotic agents (Chap. 60), initially appears similar to an exanthematous drug reaction (Fig. A1-48) but may progress to exfoliative erythroderma; it is accompanied by multiorgan failure and has an associated mortality rate of ~10%.

■ VESICULOBULLOUS OR PUSTULAR ERUPTIONS

Varicella (Chap. 193) is highly contagious, often occurring in winter or spring, and is characterized by pruritic lesions that, within a given region of the body, are in different stages of development at any point in time (Fig. 19-4; see also Fig. A1-30). In immunocompromised hosts, varicella vesicles may lack the characteristic erythematous base or may appear hemorrhagic. Lesions of *Pseudomonas* “hot-tub” folliculitis (Chap. 164) are also pruritic and may appear similar to those of varicella (Fig. A1-55). However, hot-tub folliculitis generally occurs in outbreaks after bathing in hot tubs or swimming pools, and lesions occur in regions occluded by bathing suits. Lesions of *variola* (smallpox) (Chap. S3) also appear similar to those of varicella but are



FIGURE 19-3 Confluent desquamative erythema in a patient with Staphylococcal scalded-skin syndrome. Nikolsky sign evident as shearing of epidermis due to gentle, lateral pressure. (From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 554, Figure 25-42; with permission.)



FIGURE 19-4 Vesicular and pustular lesions on the chest in a patient with varicella. (From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 695, Figure 27-48; with permission.)

all at the same stage of development in a given region of the body (**Figs. A1-50B, A1-50C**). Variola lesions are most prominent on the face (**Fig. A1-50A**) and extremities, while varicella lesions are most prominent on the trunk. *Herpes simplex virus* infection (**Chap. 192**) is characterized by hallmark grouped vesicles on an erythematous base. Primary herpes infection is accompanied by fever and toxicity, while recurrent disease is milder. *Rickettsialpox* (**Chap. 187**) is often documented in urban settings and is characterized by vesicles followed by pustules (**Figs. A1-33B, A1-33C**). It can be distinguished from varicella by an eschar at the site of the mouse-mite bite (**Fig. A1-33A**) and the papule/plaque base of each vesicle. *Acute generalized exanthematous pustulosis* (**Fig. A1-49**) should be considered in individuals who are acutely febrile and are taking new medications, especially anticonvulsant or antimicrobial agents (**Chap. 60**). Disseminated *Vibrio vulnificus* infection (**Chap. 168**) or *ecthyma gangrenosum* due to *Pseudomonas aeruginosa* (**Fig. A1-34**) (**Chap. 164**) should be considered in immunosuppressed individuals with sepsis and hemorrhagic bullae. In children, *Mycoplasma pneumoniae*-induced rash and mucositis (MIRM) (**Fig. A1-56**) is characterized by a sparse, often vesiculobullous eruption with prominent oral, ocular, or urogenital mucositis.

■ URTICARIA-LIKE ERUPTIONS

Individuals with classic urticaria ("hives") (**Fig. 19-5; see also Fig. A1-35**) usually have a hypersensitivity reaction without associated fever. In the presence of fever, urticaria-like eruptions are most often due to *urticarial vasculitis* (**Chap. 363**). Unlike individual lesions of classic urticaria, which last up to 24 h, these lesions may last 3–5 days. Etiologies include serum sickness (often induced by drugs such as penicillins, sulfas, salicylates, or barbiturates), connective-tissue disease (e.g., systemic lupus erythematosus or Sjögren's syndrome), and infection (e.g., with hepatitis B virus, enteroviruses, or parasites). Malignancy, especially lymphoma, may be associated with fever and chronic urticaria (**Chap. 58**).



FIGURE 19-5 Urticarial eruption. (From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 299, Figure 14-2; with permission.)

■ NODULAR ERUPTIONS

In immunocompromised hosts, nodular lesions often represent disseminated infection. Patients with disseminated *candidiasis* (**Fig. A1-37**) (often due to *Candida tropicalis*) may have a triad of fever, myalgias, and eruptive nodules (**Chap. 216**). Disseminated *cryptococcosis* lesions (**Fig. 19-6; see also Fig. A1-36**) (**Chap. 215**) may resemble molluscum contagiosum (**Chap. 196**). Necrosis of nodules should raise the suspicion of *aspergillosis* (**Fig. A1-38**) (**Chap. 217**) or *mucormycosis*



FIGURE 19-6 Nodular eruption on the face due to disseminated Cryptococcus in a patient with HIV infection. (From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 641, Figure 26-57. Used with permission from Loïc Vallant, MD.)



FIGURE 19-7 Purpura fulminans in a patient with acute meningococcemia. (From K Wolff et al [eds]: *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 8th ed. New York, McGraw-Hill, 2017, p. 568, Figure 25-59; with permission.)

([Chap. 218](#)). Erythema nodosum presents with exquisitely tender nodules on the lower extremities ([Fig. A1-39](#)). Sweet syndrome ([Chap. 58](#)) should be considered in individuals with multiple nodules and plaques, often so edematous ([Fig. A1-40](#)) that they give the appearance of vesicles or bullae. Sweet syndrome may occur in individuals with infection, inflammatory bowel disease, or malignancy and can also be induced by drugs.

PURPURIC ERUPTIONS

Acute meningococcemia ([Chap. 155](#)) classically presents in children as a petechial eruption, but initial lesions may appear as blanchable macules or urticaria. Rocky Mountain spotted fever should be considered in the differential diagnosis of acute meningococcemia. *Echovirus 9 infection* ([Chap. 204](#)) may mimic acute meningococcemia; patients should be treated as if they have bacterial sepsis because prompt differentiation of these conditions may be impossible. Large ecchymotic areas of *purpura*

fulminans ([Fig. 19-7](#); see also [Fig. A1-41](#)) ([Chaps. 155 and 304](#)) reflect severe underlying disseminated intravascular coagulation, which may be due to infectious or noninfectious causes. The lesions of *chronic meningococcemia* ([Fig. A1-42](#)) ([Chap. 155](#)) may have a variety of morphologies, including petechial. Purpuric nodules may develop on the legs and resemble erythema nodosum but lack its exquisite tenderness. Lesions of *disseminated gonococcemia* ([Chap. 156](#)) are distinctive, sparse, countable hemorrhagic pustules ([Fig. A1-43](#)), usually located near joints. The lesions of chronic meningococcemia and those of gonococcemia may be indistinguishable in terms of appearance and distribution. *Viral hemorrhagic fever* ([Chaps. 209 and 210](#)) should be considered in patients with an appropriate travel history and a petechial rash. *Thrombotic thrombocytopenic purpura* ([Chaps. 58, 100, and 115](#)) and *hemolytic-uremic syndrome* ([Chaps. 115, 161, and 166](#)) are closely related and are non-infectious causes of fever and petechiae. *Cutaneous small-vessel vasculitis (leukocytoclastic vasculitis)* typically manifests as palpable purpura ([Fig. A1-44](#)) and has a wide variety of causes ([Chap. 58](#)).

ERUPTIONS WITH ULCERS OR ESCARS

The presence of an ulcer or eschar ([Fig. 19-8](#)) in the setting of a more widespread eruption can provide an important diagnostic clue. For example, an eschar may suggest the diagnosis of *scrub typhus* or *rickettsialpox* ([Fig. A1-33A](#)) ([Chap. 187](#)) in the appropriate setting. In other illnesses (e.g., anthrax) ([Fig. A1-52](#)) ([Chap. S3](#)), an ulcer or eschar may be the only skin manifestation.

FURTHER READING

- CHERRY JD: Cutaneous manifestations of systemic infections, in *Feigin and Cherry's Textbook of Pediatric Infectious Diseases*, 8th ed. JD Cherry et al (eds). Philadelphia, Elsevier, 2019, pp 539–559.
- JULIANO JJ et al: The acutely ill patient with fever and rash, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, vol 1, 9th ed. JI Bennett et al (eds). Philadelphia, Elsevier, 2020, pp 801–818.
- KANG S et al (eds): *Fitzpatrick's Dermatology*, 9th ed. New York, McGraw-Hill, 2019.
- WOLFF K et al: *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 8th ed. New York, McGraw-Hill, 2017.



FIGURE 19-8 Eschar with surrounding erythema at the site of a tick bite in a patient with African tick-bite fever. (From K Wolff et al [eds]: *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 8th ed. New York, McGraw-Hill, 2017, p. 561, Figure 25-49; with permission.)

20

Fever of Unknown Origin

Chantal P. Bleeker-Rovers,
Catharina M. Mulders-Manders,
Jos W. M. van der Meer

criterion: FUO is defined as the presence of elevated inflammatory parameters (CRP or ESR) on multiple occasions for a period of at least 3 weeks in an immunocompetent patient with normal body temperature, for which a final explanation is lacking despite history-taking, physical examination, and the obligatory tests listed above. It has been shown that the causes and workup for IUO are the same as for FUO. Therefore, for convenience, the term FUO will refer to both FUO and IUO within the remainder of this chapter.

DEFINITION

Clinicians commonly refer to any febrile illness without an initially obvious etiology as *fever of unknown origin* (FUO). Most febrile illnesses either resolve before a diagnosis can be made or develop distinguishing characteristics that lead to a diagnosis. The term *FUO* should be reserved for prolonged febrile illnesses without an established etiology despite intensive evaluation and diagnostic testing. This chapter focuses on FUO in the adult patient.

FUO was originally defined by Petersdorf and Beeson in 1961 as an illness of >3 weeks' duration with fever of ≥38.3°C (≥101°F) on two occasions and an uncertain diagnosis despite 1 week of inpatient evaluation. Nowadays, most patients with FUO are hospitalized only if their clinical condition requires it, and not for diagnostic purposes alone; thus the in-hospital evaluation requirement has been eliminated from the definition. The definition of FUO has been further modified by the exclusion of immunocompromised patients, whose workup requires an entirely different diagnostic and therapeutic approach. For optimal comparison of patients with FUO in different geographic areas, it has been proposed that the quantitative criterion (diagnosis uncertain after 1 week of evaluation) be changed to a qualitative criterion that requires the performance of a specific list of investigations. Accordingly, FUO is now defined as follows:

1. Fever ≥38.3°C (≥101°F) on at least two occasions
2. Illness duration of ≥3 weeks
3. No known immunocompromised state
4. Diagnosis that remains uncertain after a thorough history-taking, physical examination, and the following obligatory investigations: determination of erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) level; platelet count; leukocyte count and differential; measurement of levels of hemoglobin, electrolytes, creatinine, total protein, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, creatine kinase, ferritin, antinuclear antibodies, and rheumatoid factor; protein electrophoresis; urinalysis; blood cultures ($n = 3$); urine culture; chest x-ray; abdominal ultrasonography; and tuberculin skin test (TST) or interferon γ release assay (IGRA).

Closely related to FUO is *inflammation of unknown origin* (IUO), which has the same definition as FUO, except for the body temperature

ETIOLOGY AND EPIDEMIOLOGY

Table 20-1 summarizes the findings of large studies on FUO conducted over the past 20 years.

The range of FUO etiologies has evolved since its first definition as a result of changes in the spectrum of diseases causing FUO, the widespread use of antibiotics, and especially the availability of new diagnostic techniques. The proportion of cases caused by intraabdominal abscesses and tumors, for example, has decreased because of earlier detection by CT and ultrasound. In addition, infective endocarditis is a less frequent cause because blood culture and echocardiographic techniques have improved. Conversely, some diagnoses such as acute HIV infection were unknown six decades ago.

Roughly comparable to 60 years ago, in non-Western cohorts infections remain the most common cause of FUO. Up to half of all infections in patients with FUO outside Western nations are caused by *Mycobacterium tuberculosis*, which is a less common cause in Western Europe and probably also in the United States. Recent data from the latter, however, have not been reported. In Western cohorts, noninfectious inflammatory diseases (NIIDs), including autoimmune, autoinflammatory, and granulomatous diseases, as well as vasculitides, form the most common cause of FUO. More than one-third of Western patients with FUO have a diagnosis that falls within the category of NIIDs. The number of FUO patients diagnosed with NIIDs probably will not decrease in the near future, as fever may precede more typical manifestations or laboratory evidence of these diseases by months. Moreover, many NIIDs can be diagnosed only after prolonged observation and exclusion of other diseases.

In Western cohorts, FUO remains unexplained in more than one-third of patients. This is much higher than 60 years ago. This difference can be explained by the fact that in patients with fever a diagnosis is often established before 3 weeks have elapsed because these patients tend to seek medical advice earlier, and because better diagnostic techniques, such as CT, MRI, and positron emission tomography (PET)/CT, are now available. Therefore, only the cases that are most difficult to diagnose continue to meet the criteria for FUO. Furthermore, most patients who have FUO without a diagnosis currently do well. A less aggressive diagnostic approach may be used in clinically stable patients once diseases with immediate therapeutic or prognostic consequences have been ruled out. In patients with recurrent fever (defined as repeated episodes of fever

TABLE 20-1 Etiology of FUO: Pooled Results of Large Studies Published in the Past 20 Years (1999–2019)

GEOGRAPHIC AREA	NO. OF COHORTS (INCLUSION PERIOD)	NO. OF PATIENTS	INFECTIONS, MEDIAN % (RANGE)	NONINFECTIOUS INFLAMMATORY DISEASES, MEDIAN % (RANGE)	MALIGNANCY, MEDIAN % (RANGE)	MISCELLANEOUS, MEDIAN % (RANGE)	NO DIAGNOSIS, MEDIAN % (RANGE)
Western Europe	10 (1990–2014)	1820	17 (11–32)	25 (12–32)	10 (3–20)	10 (0–15)	37 (26–51)
Other European and Turkey	13 (1984–2015)	1316	38 (26–59)	25 (15–38)	14 (5–19)	6 (2–18)	16 (4–35)
Middle East	3 2009–2010 and ? ^a	1235	66 (42–79)	15 (7–17)	7 (1–30)	1 (0–12)	8 (2–12)
Asia	20 (1994–2017)	3802	42 (11–58)	20 (7–57)	13 (6–22)	9 (0–15)	18 (0–36)

^aOne study (published in 2015) did not report the inclusion period.

Abbreviation: NIID, non-infectious inflammatory disease.

For references, see supplementary material at www.accessmedicine.com/harrison.

interspersed with fever-free intervals of at least 2 weeks and apparent remission of the underlying disease), the chance of attaining an etiologic diagnosis is <50%.

■ DIFFERENTIAL DIAGNOSIS

The differential diagnosis for FUO is extensive. It is important to remember that FUO is far more often caused by an atypical presentation of a rather common disease than by a very rare disease. **Table 20-2** presents an overview of possible causes of FUO. Atypical presentations of endocarditis, diverticulitis, vertebral osteomyelitis, and extrapulmonary tuberculosis are the more common infectious disease diagnoses.

Q fever and Whipple's disease (*Tropheryma whipplei* infection) are quite rare but should always be kept in mind as a cause of FUO since the presenting symptoms can be nonspecific. Serologic testing for Q fever, which results from exposure to animals or animal products, should be performed by immunofluorescence assay (IFA) when the patient lives in a rural area or has a history of heart valve disease, an aortic aneurysm, or a vascular prosthesis. In patients with unexplained symptoms localized to the central nervous system, gastrointestinal tract, or joints, polymerase chain reaction testing for *Tropheryma whipplei* should be performed. Travel to or (former) residence in tropical countries or the American Southwest should lead to consideration

TABLE 20-2 All Reported Causes of Fever of Unknown Origin (FUO)^a

Infections	
Bacterial, nonspecific	Abdominal abscess, adnexitis, apical granuloma, appendicitis, cholangitis, cholecystitis, diverticulitis, endocarditis, endometritis, epidural abscess, infected joint prosthesis, infected vascular catheter, infected vascular prosthesis, infectious arthritis, infective myonecrosis, intracranial abscess, liver abscess, lung abscess, malakoplakia, mastoiditis, mediastinitis, mycotic aneurysm, osteomyelitis, pelvic inflammatory disease, prostatitis, pyelonephritis, pylephlebitis, renal abscess, septic phlebitis, sinusitis, spondylodiscitis, xanthogranulomatous urinary tract infection
Bacterial, specific	Actinomycosis, atypical mycobacterial infection, bartonellosis, brucellosis, <i>Campylobacter</i> infection, <i>Chlamydia pneumoniae</i> infection, chronic meningococcemia, ehrlichiosis, gonococcemia, legionellosis, leptospirosis, listeriosis, louse-borne relapsing fever (<i>Borrelia recurrentis</i>), Lyme disease, melioidosis (<i>Pseudomonas pseudomallei</i>), <i>Mycoplasma</i> infection, nocardiosis, psittacosis, Q fever (<i>Coxiella burnetii</i>), rickettsiosis, <i>Spirillum minor</i> infection, <i>Streptobacillus moniliformis</i> infection, syphilis, tick-borne relapsing fever (<i>Borrelia duttonii</i>), tuberculosis, tularemia, typhoid fever and other salmonelloses, Whipple's disease (<i>Tropheryma whipplei</i>), yersiniosis
Fungal	Aspergillosis, blastomycosis, candidiasis, coccidioidomycosis, cryptococcosis, histoplasmosis, <i>Malassezia furfur</i> infection, paracoccidioidomycosis, <i>Pneumocystis jirovecii</i> pneumonia, sporotrichosis, zygomycosis
Parasitic	Amebiasis, babesiosis, echinococcosis, fascioliasis, malaria, schistosomiasis, strongyloidiasis, toxocariasis, toxoplasmosis, trichinellosis, trypanosomiasis, visceral leishmaniasis
Viral	Colorado tick fever, coxsackievirus infection, cytomegalovirus infection, dengue, Epstein-Barr virus infection, hantavirus infection, hepatitis (A, B, C, D, E), herpes simplex, HIV infection, human herpesvirus 6 infection, parvovirus infection, West Nile virus infection
Noninfectious Inflammatory Diseases	
Systemic rheumatic and autoimmune diseases	Ankylosing spondylitis, antiphospholipid syndrome, autoimmune hemolytic anemia, autoimmune hepatitis, Behcet's disease, cryoglobulinemia, dermatomyositis, Felty syndrome, gout, mixed connective-tissue disease, polymyositis, pseudogout, reactive arthritis, relapsing polychondritis, rheumatic fever, rheumatoid arthritis, Sjögren's syndrome, systemic lupus erythematosus, Vogt-Koyanagi-Harada syndrome
Vasculitis	Allergic vasculitis, eosinophilic granulomatosis with polyangiitis, giant cell vasculitis/polymyalgia rheumatica, granulomatosis with polyangiitis, hypersensitivity vasculitis, Kawasaki disease, polyarteritis nodosa, Takayasu arteritis, urticarial vasculitis
Granulomatous diseases	Idiopathic granulomatous hepatitis, sarcoidosis
Autoinflammatory syndromes	Adult-onset Still's disease, Blau syndrome, CAPS ^b (cryopyrin-associated periodic syndromes), Crohn's disease, DIRA (deficiency of the interleukin 1 receptor antagonist), familial Mediterranean fever, hemophagocytic syndrome, hyper-IgD syndrome (HIDS, also known as mevalonate kinase deficiency), juvenile idiopathic arthritis, PAPA syndrome (pyogenic sterile arthritis, pyoderma gangrenosum, and acne), PFAPA syndrome (periodic fever, aphthous stomatitis, pharyngitis, adenitis), recurrent idiopathic pericarditis, SAPHO (synovitis, acne, pustulosis, hyperostosis, osteomyelitis), Schnitzler syndrome, TRAPS (tumor necrosis factor receptor-associated periodic syndrome)
Neoplasms	
Hematologic malignancies	Amyloidosis, angioimmunoblastic lymphoma, Castleman's disease, Hodgkin's disease, hypereosinophilic syndrome, leukemia, lymphomatoid granulomatosis, malignant histiocytosis, multiple myeloma, myelodysplastic syndrome, myelofibrosis, non-Hodgkin's lymphoma, plasmacytoma, systemic mastocytosis, vaso-occlusive crisis in sickle cell disease
Solid tumors	Most solid tumors and metastases can cause fever. Those most commonly causing FUO are breast, colon, hepatocellular, lung, pancreatic, and renal cell carcinomas.
Benign tumors	Angiomyolipoma, cavernous hemangioma of the liver, craniopharyngioma, necrosis of dermoid tumor in Gardner's syndrome
Miscellaneous Causes	
	ADEM (acute disseminated encephalomyelitis), adrenal insufficiency, aneurysms, anomalous thoracic duct, aortic dissection, aortic-enteral fistula, aseptic meningitis (Mollaret's syndrome), atrial myxoma, brewer's yeast ingestion, Caroli disease, cholesterol emboli, cirrhosis, complex partial status epilepticus, cyclic neutropenia, drug fever, Erdheim-Chester disease, extrinsic allergic alveolitis, Fabry's disease, factitious disease, fire-eater's lung, fraudulent fever, Gaucher disease, Hamman-Rich syndrome (acute interstitial pneumonia), Hashimoto's encephalopathy, hematoma, hypersensitivity pneumonitis, hypertriglyceridemia, hypothalamic hypopituitarism, idiopathic normal-pressure hydrocephalus, inflammatory pseudotumor, Kikuchi's disease, linear IgA dermatosis, mesenteric fibromatosis, metal fume fever, milk protein allergy, myotonic dystrophy, nonbacterial osteitis, organic dust toxic syndrome, panniculitis, POEMS (polyneuropathy, organomegaly, endocrinopathy, monoclonal protein, skin changes), polymer fume fever, post-cardiac injury syndrome, primary biliary cirrhosis, primary hyperparathyroidism, pulmonary embolism, pyoderma gangrenosum, retroperitoneal fibrosis, Rosai-Dorfman disease, sclerosing mesenteritis, silicone embolization, subacute thyroiditis (de Quervain's), Sweet syndrome (acute febrile neutrophilic dermatosis), thrombosis, tubulointerstitial nephritis and uveitis syndrome (TINU), ulcerative colitis
Thermoregulatory Disorders	
Central	Brain tumor, cerebrovascular accident, encephalitis, hypothalamic dysfunction
Peripheral	Anhidrotic ectodermal dysplasia, exercise-induced hyperthermia, hyperthyroidism, pheochromocytoma

^aThis table includes all causes of FUO that have been described in the literature. ^bCAPS includes chronic infantile neurologic cutaneous and articular syndrome (CINCA, also known as neonatal-onset multisystem inflammatory disease, or NOMID), familial cold autoinflammatory syndrome (FCAS), and Muckle-Wells syndrome.

of infectious diseases such as malaria, leishmaniasis, histoplasmosis, or coccidioidomycosis. Fever with signs of endocarditis and negative blood culture results poses a special problem. Culture-negative endocarditis (**Chap. 128**) may be due to difficult-to-culture bacteria such as nutritionally variant bacteria, HACEK organisms (including *Haemophilus parainfluenzae*, *H. paraphrophilus*, *Aggregatibacter actinomycetemcomitans*, *A. aphrophilus*, *A. paraphrophilus*, *Cardiobacterium hominis*, *C. valvarum*, *Eikenella corrodens*, and *Kingella kingae*; discussed below), *Coxiella burnetii*, *T. whipplei*, and *Bartonella* species. Marantic endocarditis is a sterile thrombotic disease that occurs as a paraneoplastic phenomenon, especially with adenocarcinomas. Sterile endocarditis is also seen in the context of systemic lupus erythematosus and antiphospholipid syndrome.

Of the NIIDs, adult-onset Still's disease, large-vessel vasculitis, polymyalgia rheumatica, systemic lupus erythematosus (SLE), and sarcoidosis are rather common diagnoses in patients with FUO. The hereditary autoinflammatory syndromes are very rare (with the exception of familial Mediterranean fever in specific geographic regions) and usually present in young patients. Schnitzler syndrome, which can present at any age, is uncommon but can often be diagnosed easily in a patient with FUO who presents with urticaria, bone pain, and monoclonal gammopathy.

Although most tumors can present with fever, malignant lymphoma is by far the most common diagnosis of FUO among the neoplasms. Sometimes the fever even precedes lymphadenopathy detectable by physical examination.

Apart from drug-induced fever and exercise-induced hyperthermia, none of the miscellaneous causes of fever is found very frequently in patients with FUO. Virtually all drugs can cause fever, even after long-term use. *Drug-induced fever*, including DRESS (drug reaction with eosinophilia and systemic symptoms; **Fig. A1-48**), is often accompanied by eosinophilia and also by lymphadenopathy, which can be extensive. More common causes of drug-induced fever are allopurinol, carbamazepine, lamotrigine, phenytoin, sulfasalazine, furosemide, antimicrobial drugs (especially sulfonamides, minocycline, vancomycin, β -lactam antibiotics, and isoniazid), some cardiovascular drugs (e.g., quinidine), and some antiretroviral drugs (e.g., nevirapine). *Exercise-induced hyperthermia* (**Chaps. 18 and 465**) is characterized by an elevated body temperature that is associated with moderate to strenuous exercise lasting from half an hour up to several hours without an increase in CRP level or ESR. Unlike patients with fever, these patients typically sweat during the temperature elevation. *Factitious fever* (fever artificially induced by the patient—for example, by IV injection of contaminated water) should be considered in all patients but is more common among young women in health-care professions. In *fraudulent fever*, the patient is normothermic but manipulates the thermometer. Simultaneous measurements at different body sites (rectum, ear, mouth) should rapidly identify this diagnosis. Another clue to fraudulent fever is dissociation between pulse rate and temperature.

Previous studies of FUO have shown that a cause is more likely to be found in elderly patients than in younger age groups. In many cases, FUO in the elderly results from an atypical manifestation of a common disease, among which giant cell arteritis and polymyalgia rheumatica are most frequently involved. Tuberculosis is the most common infectious disease associated with FUO in elderly patients, occurring much more often than in younger patients. As many of these diseases are treatable, it is well worth pursuing the cause of fever in elderly patients.

APPROACH TO THE PATIENT

Fever of Unknown Origin

FIRST-STAGE DIAGNOSTIC TESTS

Figure 20-1 shows a structured approach to patients presenting with FUO. The most important step in the diagnostic workup is the search for potentially diagnostic clues (PDCs) through complete and repeated history-taking and physical examination and the obligatory investigations listed above and in the figure. PDCs are defined as all localizing signs, symptoms, and abnormalities

potentially pointing toward a diagnosis. Although PDCs are often misleading, only with their help can a concise list of probable diagnoses be made. The history should include information about the fever pattern (continuous or recurrent) and duration, previous medical history, present and recent drug use, family history, sexual history, country of origin, recent and remote travel, unusual environmental exposures associated with travel or hobbies, and animal contacts. A complete physical examination should be performed, with special attention to the eyes, lymph nodes, temporal arteries, liver, spleen, sites of previous surgery, entire skin surface, and mucous membranes. Before further diagnostic tests are initiated, antibiotic and glucocorticoid treatment, which can mask many diseases, should be stopped. For example, blood and other cultures are not reliable when samples are obtained during antibiotic treatment, and the size of enlarged lymph nodes usually decreases during glucocorticoid treatment, regardless of the cause of lymphadenopathy. Despite the high percentage of false-positive ultrasounds and the relatively low sensitivity of chest x-rays, the performance of these simple, low-cost diagnostic tests remains obligatory in all patients with FUO in order to separate cases that are caused by easily diagnosed diseases from those that are not. Abdominal ultrasound is preferred to abdominal CT as an obligatory test because of relatively low cost, lack of radiation burden, and absence of side effects.

Only rarely do biochemical tests (beyond the obligatory tests needed to classify a patient's fever as FUO) lead directly to a definitive diagnosis in the absence of PDCs. The diagnostic yield of immunologic serology other than that included in the obligatory tests is relatively low. These tests more often yield false-positive rather than true-positive results and are of little use without PDCs pointing to specific immunologic disorders. Given the absence of specific symptoms in many patients and the relatively low cost of the test, investigation of cryoglobulins appears to be a valuable screening test in patients with FUO.

Multiple blood samples should be cultured in the laboratory long enough to ensure ample growth time for any fastidious organisms, such as HACEK organisms. It is critical to inform the laboratory of the intent to test for unusual organisms. Specialized media should be used when the history suggests uncommon microorganisms, such as *Histoplasma* or *Legionella*. Performing more than three blood cultures or more than one urine culture is useless in patients with FUO in the absence of PDCs (e.g., a high level of clinical suspicion of endocarditis). Repeating blood or urine cultures is useful only when previously cultured samples were collected during antibiotic treatment or within 1 week after its discontinuation. FUO with headache should prompt microbiologic examination of cerebrospinal fluid (CSF) for organisms including herpes simplex virus (especially type 2), *Cryptococcus neoformans*, and *Mycobacterium tuberculosis*. In central nervous system tuberculosis, the CSF typically has elevated protein and lowered glucose concentrations, with a mononuclear pleocytosis. CSF protein levels range from 100 to 500 mg/dL in most patients, the CSF glucose concentration is <45 mg/dL in 80% of cases, and the usual CSF cell count is between 100 and 500 cells/ μ L.

Microbiologic serology should not be included in the diagnostic workup of patients without PDCs for specific infections. A tuberculin skin test (TST) or interferon γ release assay (IGRA, QuantiFERON test) is included in the obligatory investigations, but it may yield false-negative results in patients with miliary tuberculosis, malnutrition, or immunosuppression. Although the IGRA is less influenced by prior vaccination with bacille Calmette-Guérin (BCG) or by infection with nontuberculous mycobacteria, its sensitivity is similar to that of the TST; a negative TST or IGRA therefore does not exclude a diagnosis of tuberculosis. Miliary tuberculosis is especially difficult to diagnose. Granulomatous disease in liver or bone marrow biopsy samples, for example, should always lead to a (re)consideration of this diagnosis. If miliary tuberculosis is suspected, liver biopsy for acid-fast smear, culture, and polymerase chain reaction probably still has the highest diagnostic yield;

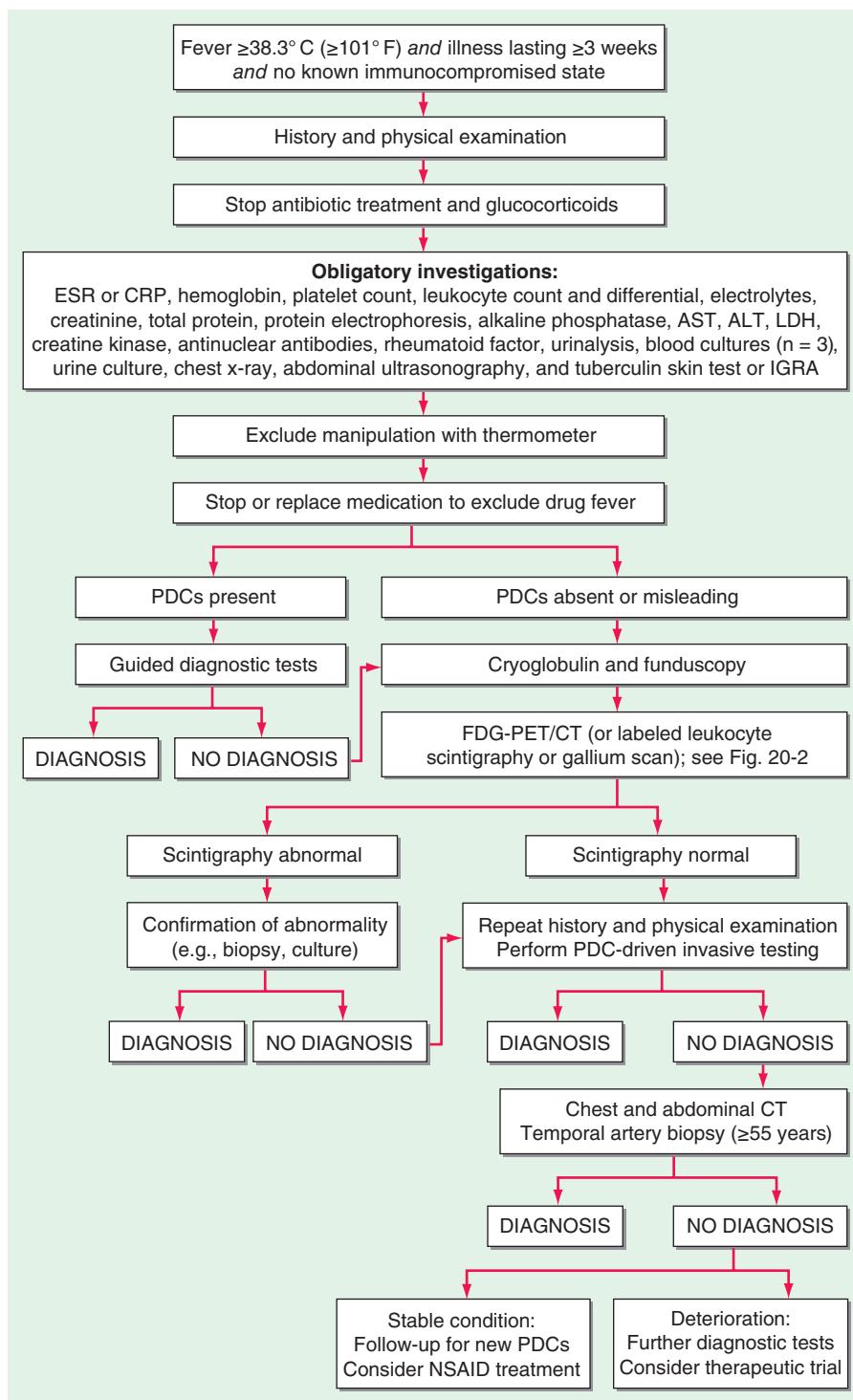


FIGURE 20-1 Structured approach to patients with FUO. ALT, alanine aminotransferase; AST, aspartate aminotransferase; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FDG-PET/CT, ¹⁸F-fluorodeoxyglucose positron emission tomography combined with low-dose CT; IGRA, interferon γ release assay; LDH, lactate dehydrogenase; NSAID, nonsteroidal anti-inflammatory drug; PDCs, potentially diagnostic clues (all localizing signs, symptoms, and abnormalities potentially pointing toward a diagnosis).

however, biopsies of bone marrow, lymph nodes, or other involved organs also can be considered.

The diagnostic yield of echocardiography, sinus radiography, radiologic or endoscopic evaluation of the gastrointestinal tract, and bronchoscopy is very low in the absence of PDCs. Therefore, these tests should not be used as screening procedures.

After identification of all PDCs retrieved from the history, physical examination, and obligatory tests, a limited list of the most probable diagnoses should be made. Since most investigations are

helpful only for patients who have PDCs for the diagnoses sought, further diagnostic procedures should be limited to specific investigations aimed at confirming or excluding diseases on this list. In FUO, the diagnostic pointers are numerous and diverse but may be missed on initial examination, often being detected only by a very careful examination performed subsequently. In the absence of PDCs, the history and physical examination should therefore be repeated regularly. One of the first steps should be to rule out factitious or fraudulent fever, particularly in patients without signs

of inflammation in laboratory tests. All medications, including nonprescription drugs and nutritional supplements, should be discontinued early in the evaluation to exclude drug fever. If fever persists beyond 72 h after discontinuation of the suspected drug, it is unlikely that this drug is the cause. In patients without PDCs or with only misleading PDCs, fundoscopy by an ophthalmologist may be useful in the early stage of the diagnostic workup to exclude retinal vasculitis. When the first-stage diagnostic tests do not lead to a diagnosis, ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG) positron emission tomography combined with computed tomography (PET/CT) or, if the former is not available, radiolabeled leukocyte scintigraphy should be performed, especially when the ESR or the CRP level is elevated.

Recurrent Fever In patients with recurrent fever, the diagnostic workup should consist of thorough history-taking, physical examination, and obligatory tests. The search for PDCs should be directed toward clues matching known recurrent syndromes (Table 20-3). Patients should be asked to return during a febrile episode so that the history, physical examination, and laboratory tests can be repeated during a symptomatic phase. Further diagnostic tests, such as PET/CT or scintigraphic imaging (see below), should be performed only during a febrile episode or when inflammatory parameters are abnormal because abnormalities may be absent between episodes. In patients with recurrent fever lasting >2 years, it is very unlikely that the fever is caused by infection or malignancy. Further diagnostic tests in that

direction should be considered only when PDCs for infections, vasculitis syndromes, or malignancy are present or when the patient's clinical condition is deteriorating.

Fluorodeoxyglucose Positron Emission Tomography ¹⁸F-FDG PET/CT has become an established imaging procedure in FUO. FDG accumulates in tissues with a high rate of glycolysis, which occurs not only in malignant cells but also in activated leukocytes and thus permits the imaging of acute and chronic inflammatory processes. Compared with conventional scintigraphy (see below), FDG-PET/CT offers the advantages of higher resolution, greater sensitivity in chronic low-grade infections, and a high degree of accuracy in the central skeleton. Furthermore, vascular uptake of FDG is increased in patients with vasculitis (Fig. 20-2). The mechanisms responsible for FDG uptake do not allow differentiation among infection, sterile inflammation, and malignancy. However, since all of these disorders are causes of FUO, FDG-PET/CT can be used to guide additional diagnostic tests (e.g., targeted biopsies) that may yield the final diagnosis. It is important to realize that physiologic uptake of FDG may obscure pathologic foci in the brain, heart, bowel, kidneys, and bladder. FDG uptake in the heart, which obscures endocarditis, may be prevented by consumption of a low-carbohydrate diet before the PET investigation. In patients with fever, bone marrow uptake is frequently increased in a non-specific way due to cytokine activation, which upregulates glucose transporters in bone marrow cells.

TABLE 20-3 All Reported Causes of Recurrent Fever^a

Infections	
Bacterial, nonspecific	Apical granuloma, diverticulitis, prostatitis, recurrent bacteremia caused by colonic neoplasia or persistent focal infection, recurrent cellulitis, recurrent cholangitis or cholecystitis, recurrent pneumonia, recurrent sinusitis, recurrent urinary tract infection
Bacterial, specific	Bartonellosis, brucellosis, chronic gonococcemia, chronic meningococcemia, louse-borne relapsing fever (<i>Borrelia recurrentis</i>), melioidosis (<i>Pseudomonas pseudomallei</i>), Q fever (<i>Coxiella burnetii</i>), salmonellosis, <i>Spirillum minor</i> infection, <i>Streptobacillus moniliformis</i> infection, syphilis, tick-borne relapsing fever (<i>Borrelia duttonii</i>), tularemia, Whipple's disease (<i>Tropheryma whipplei</i>), yersiniosis
Fungal	Coccidioidomycosis, histoplasmosis, paracoccidioidomycosis
Parasitic	Babesiosis, malaria, toxoplasmosis, trypanosomiasis, visceral leishmaniasis
Viral	Cytomegalovirus infection, Epstein-Barr virus infection, herpes simplex
Noninfectious Inflammatory Diseases	
Systemic rheumatic and autoimmune diseases	Ankylosing spondylitis, antiphospholipid syndrome, autoimmune hemolytic anemia, autoimmune hepatitis, Behcet's disease, cryoglobulinemia, gout, polymyositis, pseudogout, reactive arthritis, relapsing polychondritis, systemic lupus erythematosus
Vasculitis	Churg-Strauss syndrome, giant cell vasculitis/polymyalgia rheumatica, hypersensitivity vasculitis, polyarteritis nodosa, urticarial vasculitis
Granulomatous diseases	Idiopathic granulomatous hepatitis, sarcoidosis
Autoinflammatory syndromes	Adult-onset Still's disease, Blau syndrome, CANDLE (chronic atypical neutrophilic dermatitis with lipodystrophy and elevated temperature syndrome), CAPS ^b (cryopyrin-associated periodic syndrome), CRMO (chronic recurrent multifocal osteomyelitis), Crohn's disease, DIRA (deficiency of the interleukin 1 receptor antagonist), familial Mediterranean fever, hemophagocytic syndrome, hyper-IgD syndrome (HIDS, also known as mevalonate kinase deficiency), juvenile idiopathic arthritis, NLRC4-activating mutations, PAPA syndrome (pyogenic sterile arthritis, pyoderma gangrenosum, and acne), PFAPA syndrome (periodic fever, aphthous stomatitis, pharyngitis, adenitis), recurrent idiopathic pericarditis, SAPHO (synovitis, acne, pustulosis, hyperostosis, osteomyelitis), SAVI (stimulator of interferon genes [STING]-associated vasculopathy with onset in infancy), Schnitzler syndrome, TRAPS (tumor necrosis factor receptor-associated periodic syndrome)
Neoplasms	
	Angioimmunoblastic lymphoma, Castleman's disease, colon carcinoma, craniopharyngioma, Hodgkin's disease, malignant histiocytosis, mesothelioma, non-Hodgkin's lymphoma
Miscellaneous Causes	
	Adrenal insufficiency, aortic-enteral fistula, aseptic meningitis (Mollaret's syndrome), atrial myxoma, brewer's yeast ingestion, cholesterol emboli, cyclic neutropenia, drug fever, extrinsic allergic alveolitis, Fabry's disease, factitious disease, fraudulent fever, Gaucher disease, hypersensitivity pneumonitis, hypertriglyceridemia, hypothalamic hypopituitarism, inflammatory pseudotumor, metal fume fever, milk protein allergy, polymer fume fever, pulmonary embolism, sclerosing mesenteritis
Thermoregulatory Disorders	
Central	Hypothalamic dysfunction
Peripheral	Anhidrotic ectodermal dysplasia, exercise-induced hyperthermia, pheochromocytoma

^aThis table includes all causes of recurrent fever that have been described in the literature. ^bCAPS includes chronic infantile neurologic cutaneous and articular syndrome (CINCA, also known as neonatal-onset multisystem inflammatory disease, or NOMID), familial cold autoinflammatory syndrome (FCAS), and Muckle-Wells syndrome.

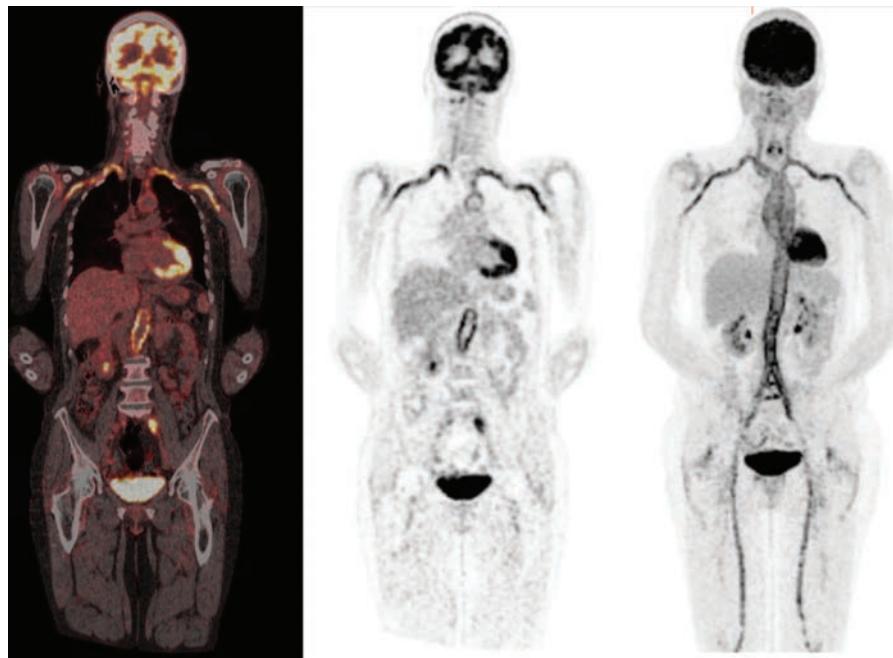


FIGURE 20-2 FDG-PET/CT in a patient with FUO. This 72-year-old woman presented with a low-grade fever and severe fatigue of almost 3 months' duration. An extensive history was taken, but the patient had no specific complaints and had not traveled recently. Her previous history was unremarkable, and she did not use any drugs. Physical examination, including palpation of the temporal arteries, yielded completely normal results. Laboratory examination showed normocytic anemia, a C-reactive protein level of 43 mg/L, an erythrocyte sedimentation rate of 87 mm/h, and mild hypoalbuminemia. Results of the other obligatory tests were all normal. Since there were no potentially diagnostic clues, FDG-PET/CT was performed. This test showed increased FDG uptake in all major arteries (carotid, jugular, and subclavian arteries; thoracic and abdominal aorta; iliac, femoral, and popliteal arteries) and in the soft tissue around the shoulders, hips, and knees—findings compatible with large-vessel vasculitis and polymyalgia rheumatica. Within 1 week after the initiation of treatment with prednisone (60 mg once daily), the patient completely recovered. After 1 month, the prednisone dose was slowly tapered.

In recent years, many cohort studies and several meta-analyses have focused on the diagnostic yield of PET and PET/CT in FUO. These studies are highly variable in terms of the selection of patients, the follow-up, and the selection of a gold-standard reference. Indirect comparisons of test performance suggested that FDG-PET/CT outperformed stand-alone FDG-PET, gallium scintigraphy, and leukocyte scintigraphy. Similarly, indirect comparisons of diagnostic yield suggested that FDG-PET/CT was more likely than alternative tests to correctly identify the cause of FUO. Meta-analyses report a high diagnostic yield for PET and PET/CT in the workup of FUO patients, with pooled sensitivity and specificity figures of ~85% and ~50%, respectively, and a total diagnostic yield of ~50% for PET/CT and ~40% for PET.

As many patients with FUO present with periodic fever, correct timing of PET/CT increases its diagnostic value. Few studies on the use of biomarkers such as elevated CRP or ESR for a contributory outcome of PET/CT have been performed. When both CRP and ESR are normal at the time of FDG-PET/CT, outcome may only be contributory when a patient does have fever at the time of the scan.

Although PET/CT and other scintigraphic techniques do not directly provide a definitive diagnosis (with the exception of some patients with, for instance, large vessel vasculitis), they often identify the anatomic location of a particular ongoing metabolic process. With the help of other techniques such as biopsy and culture, a timely diagnosis and treatment can be facilitated. Pathologic FDG uptake is quickly eradicated by treatment with glucocorticoids in many diseases, including vasculitis and lymphoma; therefore, glucocorticoid use should be stopped or postponed until after FDG-PET/CT is performed.

FDG-PET/CT is a relatively expensive procedure whose availability is still limited compared with that of CT and conventional scintigraphy. Nevertheless, FDG-PET/CT can be cost-effective in the FUO diagnostic workup if used at an early stage, helping to

establish an early diagnosis, reducing days of hospitalization for diagnostic purposes, and obviating unnecessary and unhelpful tests. When FDG-PET/CT has been made under the right conditions (i.e., when elevated CRP or ESR or fever were present during the scan) but has not contributed to the final diagnosis, repeating PET/CT is probably of little value, unless new signs or symptoms appear.

Conventional scintigraphic imaging other than PET/CT

Conventional scintigraphic methods used in clinical practice are ^{67}Ga -citrate scintigraphy and ^{111}In - or $^{99\text{m}}\text{Tc}$ -labeled leukocyte scintigraphy. Sensitivity and specificity of conventional scintigraphic studies are lower than for PET/CT: the diagnostic yield of gallium scintigraphy ranges from 21% to 54%, and on average the location of a source of fever can correctly be localized in approximately one-third of patients. The diagnostic value of leukocyte scintigraphy ranges from 8% to 31%, and overall the cause of FUO can correctly be identified in one-fifth of patients. When PET/CT is not available, these techniques are the only alternative.

LATER-STAGE DIAGNOSTIC TESTS

In some cases, more invasive tests are appropriate. Abnormalities found with imaging often need to be confirmed by pathology and/or culture of biopsy specimens. If lymphadenopathy is found, lymph node biopsy is necessary, even when the affected lymph nodes are hard to reach or when previous biopsies were inconclusive. In the case of skin lesions, skin biopsy should be undertaken.

If no diagnosis is reached despite PET/CT and PDC-driven histologic investigations or culture, second-stage screening diagnostic tests should be considered (Fig. 20-1). In three studies, the diagnostic yield of screening chest and abdominal CT in patients with FUO was ~20%. The specificity of chest CT was ~80%, but that of abdominal CT varied between 63% and 80%. Despite the

relatively limited specificity of abdominal CT and the probably limited additional value of chest CT after normal FDG-PET/CT, chest and abdominal CT may be used as screening procedures at a later stage of the diagnostic protocol because of their noninvasive nature and high sensitivity. Bone marrow aspiration is seldom useful in the absence of PDCs for bone marrow disorders. With addition of FDG-PET/CT, which is highly sensitive in detecting lymphoma, carcinoma, and osteomyelitis, the value of bone marrow biopsy as a screening procedure is probably further reduced. Several studies have shown a high prevalence of giant cell arteritis among patients with FUO, with rates up to 17% among elderly patients. Giant cell arteritis often involves large arteries and in most cases can be diagnosed by FDG-PET/CT. However, temporal artery biopsy is still recommended for patients ≥55 years of age in a later stage of the diagnostic protocol. FDG-PET/CT will not be useful in vasculitis limited to the temporal arteries because of the small diameter of these vessels and the high levels of FDG uptake in the brain. In the past, liver biopsies were often performed as a screening procedure in patients with FUO. In each of two studies, liver biopsy as part of the later stage of a screening diagnostic protocol was helpful in only one patient. Moreover, abnormal liver tests are not predictive of a diagnostic liver biopsy in FUO. Liver biopsy is an invasive procedure that carries the possibility of complications and even death. Therefore, it should not be used for screening purposes in patients with FUO except in those with PDCs for liver disease or miliary tuberculosis.

In patients with unexplained fever after all of the above procedures, the last steps in the diagnostic workup—with only a marginal diagnostic yield—come at an extraordinarily high cost in terms of both expense and discomfort for the patient. Repetition of a thorough history-taking and physical examination and review of laboratory results and imaging studies (including those from other hospitals) are recommended. Diagnostic delay often results from a failure to recognize PDCs in the available information. In these patients with persisting FUO, waiting for new PDCs to appear probably is better than ordering more screening investigations. Only when a patient's condition deteriorates without providing new PDCs should a further diagnostic workup be performed.

SECOND OPINION IN AN EXPERT CENTER

When despite the workup described above no explanation for FUO is found, second opinion in an expert center on FUO should be considered. The single study on the value of second opinion in FUO reported that in 57.3% of patients with unexplained FUO, a diagnosis could be found in an expert center. Additionally, of all patients who remained without a diagnosis even after second opinion, 10.9% became fever-free upon empirical treatment, adding up to a beneficial outcome in 68.2% of patients.

TREATMENT

Fever of Unknown Origin

Empirical therapeutic trials with antibiotics, glucocorticoids, or antituberculous agents should be avoided in FUO except when a patient's condition is rapidly deteriorating after the aforementioned diagnostic tests have failed to provide a definite diagnosis.

ANTIBIOTICS AND ANTITUBERCULOUS THERAPY

Antibiotic or antituberculous therapy may irrevocably diminish the ability to culture fastidious bacteria or mycobacteria. However, hemodynamic instability or neutropenia is a good indication for empirical antibiotic therapy. If the TST or IGRA is positive or if granulomatous disease is present with anergy and sarcoidosis seems unlikely, a trial of therapy for tuberculosis should be started. Especially in miliary tuberculosis, it may be very difficult to obtain

a rapid diagnosis. If the fever does not respond after 6 weeks of empirical antituberculous treatment, another diagnosis should be considered.

COLCHICINE, NONSTEROIDAL ANTI-INFLAMMATORY DRUGS, AND GLUCOCORTICOIDS

Colchicine is highly effective in preventing attacks of familial Mediterranean fever (FMF) but is not always effective once an attack is well under way. When FMF is suspected, the response to colchicine is not a completely reliable diagnostic tool in the acute phase, but with colchicine treatment most patients show remarkable improvements in the frequency and severity of subsequent febrile episodes within weeks to months. Therefore, colchicine may be tried in patients with features compatible with FMF, especially when these patients originate from a high-prevalence region.

If the fever persists and the source remains elusive after completion of the later-stage investigations, supportive treatment with nonsteroidal anti-inflammatory drugs (NSAIDs) can be helpful. The response of adult-onset Still's disease to NSAIDs is dramatic in some cases.

The effects of glucocorticoids on giant cell arteritis and polymyalgia rheumatica are equally impressive. Early empirical trials with glucocorticoids, however, decrease the chances of reaching a diagnosis for which more specific and sometimes life-saving treatment might be more appropriate, such as malignant lymphoma. The ability of NSAIDs and glucocorticoids to mask fever while permitting the spread of infection or lymphoma dictates that their use should be avoided unless infectious diseases and malignant lymphoma have been largely ruled out and inflammatory disease is probable and is likely to be debilitating or threatening.

INTERLEUKIN 1 INHIBITION

Interleukin (IL) 1 is a key cytokine in local and systemic inflammation and the febrile response. The availability of specific IL-1-targeting agents has revealed a pathologic role of IL-1-mediated inflammation in a growing list of diseases. Anakinra, a recombinant form of the naturally occurring IL-1 receptor antagonist (IL-1Ra), blocks the activity of both IL-1 α and IL-1 β . Anakinra is extremely effective in the treatment of many autoinflammatory syndromes, such as FMF, cryopyrin-associated periodic syndrome, tumor necrosis factor receptor-associated periodic syndrome, mevalonate kinase deficiency (hyper IgD syndrome), Schnitzler syndrome, and adult onset Still's disease. There are many other chronic inflammatory disorders in which anti-IL-1 therapy is highly effective. A therapeutic trial with anakinra can be considered in patients whose FUO has not been diagnosed after later-stage diagnostic tests. Although most chronic inflammatory conditions without a known basis can be controlled with glucocorticoids, monotherapy with IL-1 blockade can provide improved control without the metabolic, immunologic, and gastrointestinal side effects of glucocorticoid administration.

PROGNOSIS

In patients in whom FUO remains unexplained, prognosis is favorable. Two large studies on mortality in these patients have been performed. The first study included 436 patients of whom 168 remained without a diagnosis. Of these, 4 (2.4%) died during follow-up. All 4 patients died during the index admission, and in 2 of them a diagnosis was made upon autopsy (1 had intravascular lymphoma and 1 had bilateral pneumonia). The second study included 131 patients with unexplained FUO. Of these patients, 9 (6.9%) died during a median follow-up of 5 years. In 6 of these patients the cause of death was known, and in 5 of them death was considered unrelated to the febrile disease. Overall, FUO-related mortality rates have continuously declined over recent decades. The majority of fevers are caused by treatable diseases, and the risk of death related to FUO is, of course, dependent on the underlying disease.

FURTHER READING

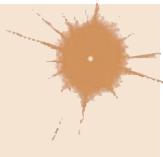
- BLEEKER-ROVERS CP et al: A prospective multicenter study on fever of unknown origin: The yield of a structured diagnostic protocol. *Medicine (Baltimore)* 86:26, 2007.
- KOIJZER IJE et al: Fever of unknown origin: The value of FDG-PET/CT. *Semin Nucl Med* 48:100, 2018.
- MULDERS-MANDERS C et al: Fever of unknown origin. *Clin Med* 15:280, 2015.
- MULDERS-MANDERS C et al: Long-term prognosis, treatment, and outcome of patients with fever of unknown origin in whom no diagnosis was made despite extensive investigation: A questionnaire based study. *Medicine (Baltimore)* 97:e11241, 2018.
- VANDERSCHUEREN S et al: Inflammation of unknown origin versus fever of unknown origin: Two of a kind. *Eur J Intern Med* 20:4, 2009.
- VANDERSCHUEREN S et al: Mortality in patients presenting with fever of unknown origin. *Acta Clin Belg* 69:12, 2014.

Section 3 Nervous System Dysfunction

21

Syncope

Roy Freeman



Syncope is a transient, self-limited loss of consciousness due to acute global impairment of cerebral blood flow. The onset is rapid, duration brief, and recovery spontaneous and complete. Other causes of transient loss of consciousness need to be distinguished from syncope; these include seizures, vertebrobasilar ischemia, hypoxemia, and hypoglycemia. A syncopal prodrome (*presyncope*) is common, although loss of consciousness may occur without any warning symptoms. Typical presyncopal symptoms include lightheadedness or faintness, dizziness, weakness, fatigue, and visual and auditory disturbances. The causes of syncope can be divided into three general categories: (1) neurally mediated syncope (also called *reflex or vasovagal syncope*), (2) orthostatic hypotension, and (3) cardiac syncope.

Neurally mediated syncope comprises a heterogeneous group of functional disorders that are characterized by a transient change in the reflexes responsible for maintaining cardiovascular homeostasis. Episodic vasodilation (or loss of vasoconstrictor tone), decreased cardiac output, and bradycardia occur in varying combinations, resulting in temporary failure of blood pressure control. In contrast, in patients with orthostatic hypotension due to autonomic failure, these cardiovascular homeostatic reflexes are chronically impaired. Cardiac syncope may be due to arrhythmias or structural cardiac diseases that cause a decrease in cardiac output. The clinical features, underlying pathophysiological mechanisms, therapeutic interventions, and prognoses differ markedly among these three causes.

EPIDEMIOLOGY AND NATURAL HISTORY

Syncope is a common presenting problem, accounting for ~3% of all emergency department (ED) visits and 1% of all hospital admissions. The annual cost for syncope-related hospitalization in the United States is ~\$2.4 billion. Syncope has a lifetime cumulative incidence of up to 35% in the general population. The peak incidence in the young occurs between ages 10 and 30 years, with a median peak around 15 years. Neurally mediated syncope is the etiology in the vast majority of these cases. In older adults, there is a sharp rise in the incidence of syncope after 70 years of age.

In population-based studies, neurally mediated syncope is the most common cause of syncope. The incidence is higher in women than men. In young subjects, there is often a family history in first-degree relatives. Cardiovascular disease due to structural disease or arrhythmias is the next most common cause in most series, particularly in ED

TABLE 21-1 High-Risk Features Indicating Hospitalization or Intensive Evaluation of Syncope

Chest pain suggesting coronary ischemia
Features of congestive heart failure
Moderate or severe valvular disease
Moderate or severe structural cardiac disease
Electrocardiographic features of ischemia
History of ventricular arrhythmias
Prolonged QT interval (>500 ms)
Repetitive sinoatrial block or sinus pauses
Persistent sinus bradycardia
Bi- or trifascicular block or intraventricular conduction delay with QRS duration ≥120 ms
Atrial fibrillation
Nonsustained ventricular tachycardia
Family history of sudden death
Preexcitation syndromes
Brugada pattern on ECG
Palpitations at time of syncope
Syncope at rest or during exercise

settings and in older patients. Orthostatic hypotension also increases in prevalence with age because of the reduced baroreflex responsiveness, decreased cardiac compliance, and attenuation of the vestibulosympathetic reflex associated with aging. Other contributors are reduced fluid intake and vasoactive medications, also more likely in this age group. In the elderly, orthostatic hypotension is more common in institutionalized than community-dwelling individuals, most likely explained by a greater prevalence of predisposing neurologic disorders, physiologic impairment, and vasoactive medication use among institutionalized patients.

Syncope of noncardiac and unexplained origin in younger individuals has an excellent prognosis; life expectancy is unaffected. By contrast, syncope due to a cardiac cause, either structural heart disease or a primary arrhythmic disorder, is associated with an increased risk of sudden cardiac death and mortality from other causes. Similarly, the mortality rate is increased in individuals with syncope due to orthostatic hypotension related to age and the associated comorbid conditions (Table 21-1). The likelihood of hospitalization and mortality risk are higher in older adults.

PATHOPHYSIOLOGY

The upright posture imposes a unique physiologic stress upon humans; most, although not all, syncopal episodes occur from a standing position. Standing results in pooling of 500–1000 mL of blood in the lower extremities, buttocks, and splanchnic circulation. The dependent pooling leads to a decrease in venous return to the heart and reduced ventricular filling that result in diminished cardiac output and blood pressure. These hemodynamic changes provoke a compensatory reflex response, initiated by the baroreceptors in the carotid sinus and aortic arch, resulting in increased sympathetic outflow and decreased vagal nerve activity (Fig. 21-1). The reflex increases peripheral resistance, venous return to the heart, and cardiac output and thus limits the fall in blood pressure. If this response fails, as is the case chronically in orthostatic hypotension and transiently in neurally mediated syncope, hypotension and cerebral hypoperfusion occur.

Syncope is a consequence of global cerebral hypoperfusion and thus represents a failure of cerebral blood flow autoregulatory mechanisms. Myogenic factors, local metabolites, and to a lesser extent autonomic neurovascular control are responsible for the autoregulation of cerebral blood flow (Chap. 307). The latency of the autoregulatory response is 5–10 s. Typically, cerebral blood flow ranges from 50–60 mL/min per 100 g brain tissue and remains relatively constant over perfusion pressures ranging from 50–150 mmHg. Cessation of blood flow for 6–8 s

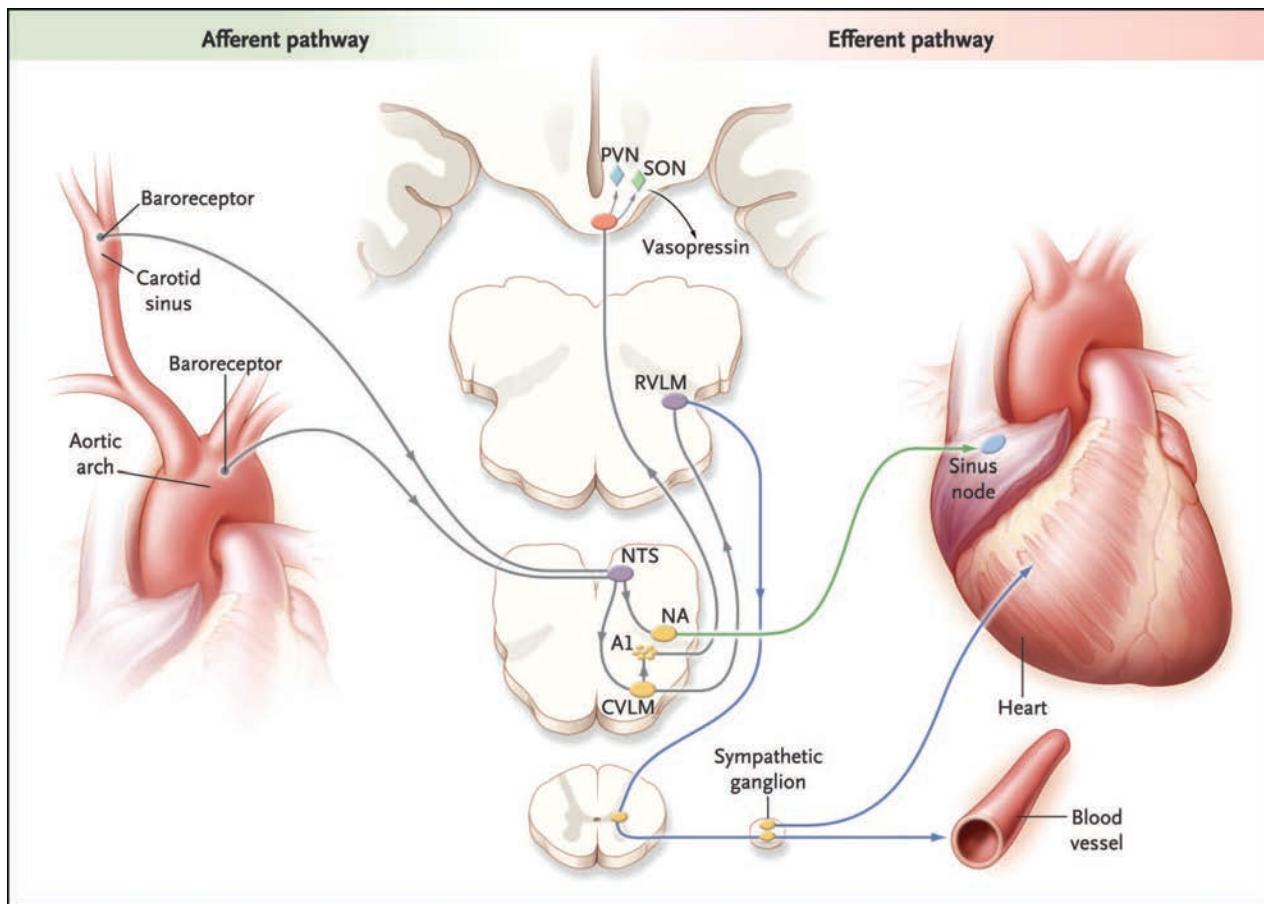


FIGURE 21-1 The baroreflex. A decrease in arterial pressure unloads the baroreceptors—the terminals of afferent fibers of the glossopharyngeal and vagus nerves—that are situated in the carotid sinus and aortic arch. This leads to a reduction in the afferent impulses that are relayed from these mechanoreceptors through the glossopharyngeal and vagus nerves to the nucleus of the tractus solitarius (NTS) in the dorsomedial medulla. The reduced baroreceptor afferent activity produces a decrease in vagal nerve input to the sinus node that is mediated via connections of the NTS to the nucleus ambiguus (NA). There is an increase in sympathetic efferent activity that is mediated by the NTS projections to the caudal ventrolateral medulla (CVLM) (an excitatory pathway) and from there to the rostral ventrolateral medulla (RVLM) (an inhibitory pathway). The activation of RVLM presynaptic neurons in response to hypotension is thus predominantly due to disinhibition. In response to a sustained fall in blood pressure, vasopressin release is mediated by projections from the A1 noradrenergic cell group in the ventrolateral medulla. This projection activates vasopressin-synthesizing neurons in the magnocellular portion of the paraventricular nucleus (PVN) and the supraoptic nucleus (SON) of the hypothalamus. Blue denotes sympathetic neurons, and green denotes parasympathetic neurons. (From R Freeman: Neurogenic orthostatic hypotension. *N Engl J Med* 358:615, 2008. Copyright © 2008 Massachusetts Medical Society. Reprinted with permission.)

will result in loss of consciousness, while impairment of consciousness ensues when blood flow decreases to 25 mL/min per 100 g brain tissue.

From the clinical standpoint, a fall in systemic systolic blood pressure to ~50 mmHg or lower will result in syncope. A decrease in cardiac output and/or systemic vascular resistance—the determinants of blood pressure—thus underlies the pathophysiology of syncope. Common causes of impaired cardiac output include decreased effective circulating blood volume, increased thoracic pressure, massive pulmonary embolus, cardiac brady- and tachyarrhythmias, valvular heart disease, and myocardial dysfunction. Systemic vascular resistance may be decreased by central and peripheral autonomic nervous system diseases, sympatholytic medications, and transiently during neurally mediated syncope. Increased cerebral vascular resistance, most frequently due to hypocapnia induced by hyperventilation, may also contribute to the pathophysiology of syncope.

Two patterns of electroencephalographic (EEG) changes occur in syncopal subjects. The first is a “slow-flat-slow” pattern (Fig. 21-2) in which normal background activity is replaced with high-amplitude slow delta waves. This is followed by sudden flattening of the EEG—a cessation or attenuation of cortical activity—followed by the return of slow waves, and then normal activity. A second pattern, the “slow pattern,” is characterized by increasing and decreasing slow wave activity only. The EEG flattening that occurs in the slow-flat-slow pattern is a marker of more severe cerebral hypoperfusion. Despite the presence of myoclonic movements and other motor activity during some syncopal events, EEG seizure discharges are not detected.

CLASSIFICATION

■ NEURALLY MEDIATED SYNCOPE

Neurally mediated (reflex; vasovagal) syncope is the final pathway of a complex central and peripheral nervous system reflex arc. There is a transient change in autonomic efferent activity with increased parasympathetic outflow, plus sympathoinhibition, resulting in bradycardia, vasodilation, and/or reduced vasoconstrictor tone (the vasodepressor response) and reduced cardiac output. The resulting fall in systemic blood pressure can then reduce cerebral blood flow to below the compensatory limits of autoregulation (Fig. 21-3). In order to develop neurally mediated syncope, a functioning autonomic nervous system is necessary, in contrast to syncope resulting from autonomic failure (discussed below).

Multiple triggers of the afferent limb of the reflex arc can result in neurally mediated syncope. In some situations, these can be clearly defined, e.g., orthostatic stress and stimulus of the carotid sinus, the gastrointestinal tract, or the bladder. Often, however, the trigger is less easily recognized and the cause is multifactorial. Under these circumstances, it is likely that different afferent pathways converge on the central autonomic network within the medulla that integrates the neural impulses and mediates the vasodepressor-bradycardic response.

Classification of Neurally Mediated Syncope Neurally mediated syncope may be subdivided based on the afferent pathway and provocative trigger. Vasovagal syncope (the common faint) is provoked

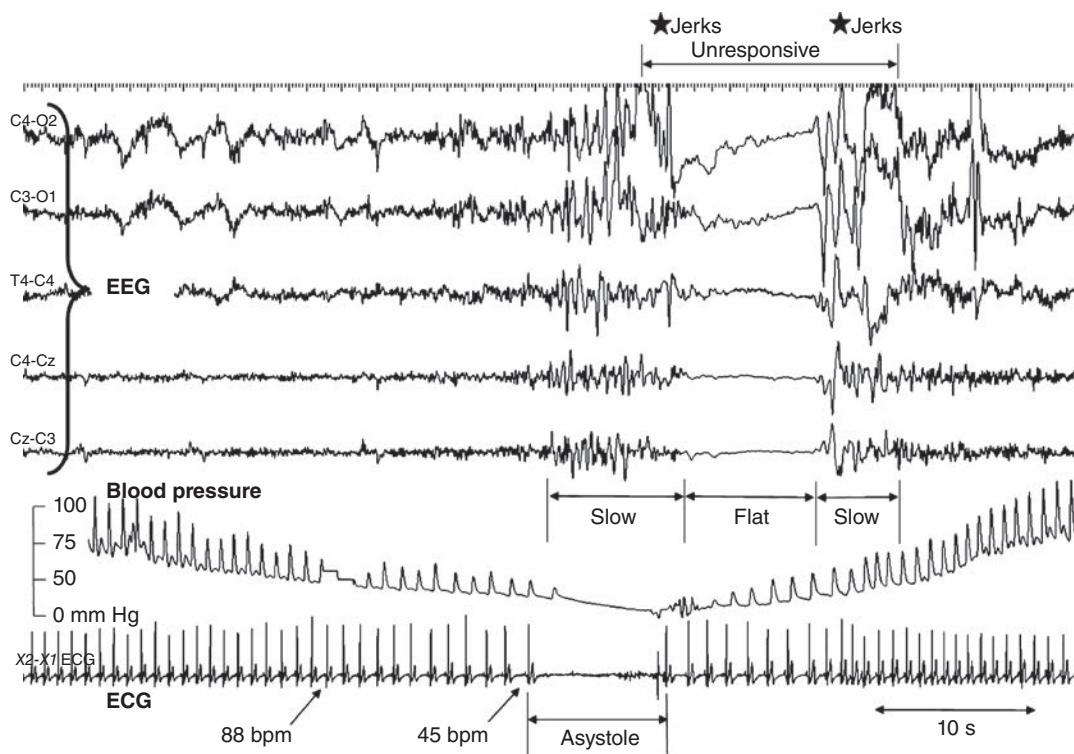


FIGURE 21-2 The electroencephalogram (EEG) in vasovagal syncope. A 1-min segment of a tilt-table test with typical vasovagal syncope demonstrating the “slow-flat-slow” EEG pattern. Finger beat-to-beat blood pressure, electrocardiogram (ECG), and selected EEG channels are shown. EEG slowing starts when systolic blood pressure drops to ~50 mmHg; heart rate is then ~45 beats/min (bpm). Asystole occurred, lasting about 8 s. The EEG flattens for a similar period, but with a delay. A transient loss of consciousness, lasting 14 s, was observed. There were muscle jerks just before and just after the flat period of the EEG. (From W Wieling et al: Symptoms and signs of syncope: a review of the link between physiology and clinical clues. *Brain* 132:2630, 2009. Reprinted (and translated) by permission of Oxford University Press on behalf of the Guarantors of Brain.)

by intense emotion, pain, and/or orthostatic stress, whereas the situational reflex syncopes have specific localized stimuli that provoke the reflex vasodilation and bradycardia that leads to syncope. The underlying mechanisms have been identified and pathophysiology delineated for most of these situational reflex syncopes. The afferent trigger may originate in the pulmonary system, gastrointestinal system, urogenital system, heart, and carotid sinus in the carotid artery (Table 21-2). Hyperventilation leading to hypocapnia and cerebral vasoconstriction, and raised intrathoracic pressure that impairs venous return to the

heart, play a central role in many of the situational reflex syncopes. The afferent pathway of the reflex arc differs among these disorders, but the efferent response via the vagus and sympathetic pathways is similar.

Alternately, neurally mediated syncope may be subdivided based on the predominant efferent pathway. Vasodepressor syncope describes syncope predominantly due to efferent, sympathetic, vasoconstrictor failure; cardioinhibitory syncope describes syncope predominantly associated with bradycardia or asystole due to increased vagal outflow;

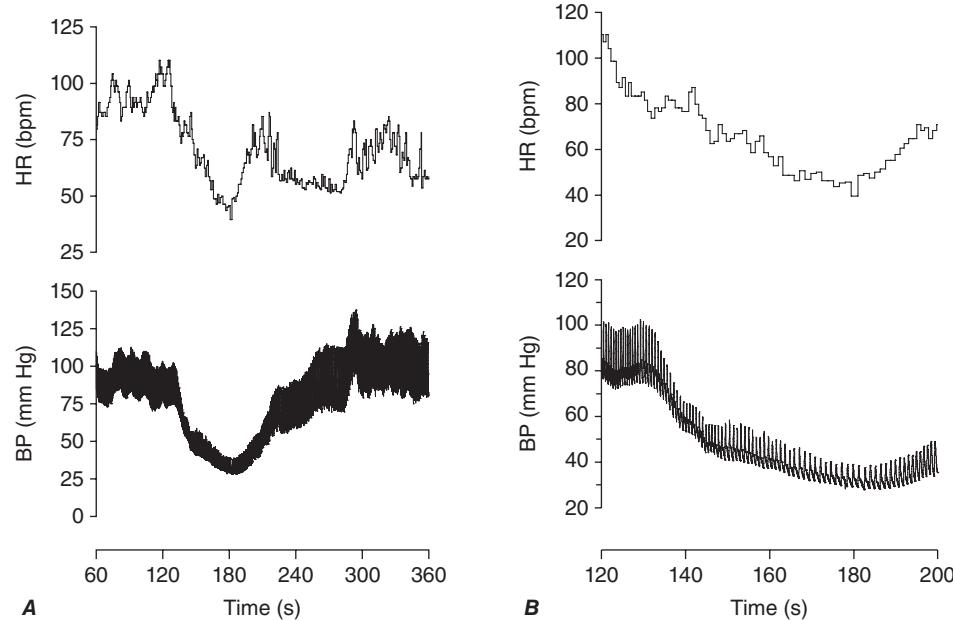


FIGURE 21-3 A. The paroxysmal hypotensive-bradycardic response that is characteristic of neurally mediated syncope. Noninvasive beat-to-beat blood pressure and heart rate are shown >5 min (from 60 to 360 s) of an upright tilt on a tilt table. B. The same tracing expanded to show 80 s of the episode (from 80 to 200 s). BP, blood pressure; bpm, beats per minute; HR, heart rate.

TABLE 21-2 Causes of Syncope**A. Neurally Mediated Syncope**

Vasovagal syncope
Provoked fear, pain, anxiety, intense emotion, sight of blood, unpleasant sights and odors, orthostatic stress
Situational reflex syncope
Pulmonary
Cough syncope, wind instrument player's syncope, weightlifter's syncope, "mess trick" ^a and "fainting lark," sneeze syncope, airway instrumentation
Urogenital
Postmicturition syncope, urogenital tract instrumentation, prostatic massage
Gastrointestinal
Swallow syncope, glossopharyngeal neuralgia, esophageal stimulation, gastrointestinal tract instrumentation, rectal examination, defecation syncope
Cardiac
Bezold-Jarisch reflex, cardiac outflow obstruction
Carotid sinus
Carotid sinus sensitivity, carotid sinus massage
Ocular
Ocular pressure, ocular examination, ocular surgery

B. Orthostatic Hypotension

Primary autonomic failure due to idiopathic central and peripheral neurodegenerative diseases—the "synucleinopathies"

Lewy body diseases
Parkinson's disease
Lewy body dementia
Pure autonomic failure
Multiple system atrophy (Shy-Drager syndrome)
Secondary autonomic failure due to autonomic peripheral neuropathies
Diabetes
Hereditary amyloidosis (familial amyloid polyneuropathy)
Primary amyloidosis (AL amyloidosis; immunoglobulin light chain associated)
Hereditary sensory and autonomic neuropathies (HSAN) (especially type III—familial dysautonomia)
Idiopathic immune-mediated autonomic neuropathy
Autoimmune autonomic ganglionopathy
Sjögren's syndrome
Paraneoplastic autonomic neuropathy
HIV neuropathy
Postprandial hypotension
Iatrogenic (drug-induced)
Volume depletion

C. Cardiac Syncope

Arrhythmias
Sinus node dysfunction
Atrioventricular dysfunction
Supraventricular tachycardias
Ventricular tachycardias
Inherited channelopathies
Cardiac structural disease
Valvular disease
Myocardial ischemia
Obstructive and other cardiomyopathies
Atrial myxoma
Pericardial effusions and tamponade

^aHyperventilation for ~1 min, followed by sudden chest compression. ^bHyperventilation (~20 breaths) in a squatting position, rapid rise to standing, then Valsalva maneuver.

and mixed syncope describes syncope in which there are both vagal and sympathetic reflex changes.

Features of Neurally Mediated Syncope In addition to symptoms of orthostatic intolerance such as dizziness, lightheadedness, and fatigue, premonitory features of autonomic activation may be present in patients with neurally mediated syncope. These include diaphoresis, pallor, palpitations, nausea, hyperventilation, and yawning. During the syncopal event, proximal and distal myoclonus (typically arrhythmic and multifocal) may occur, raising the possibility of a seizure. The eyes typically remain open and usually deviate upward. Pupils are usually dilated. Roving eye movements may occur. Grunting, moaning, snorting, and stertorous breathing may be present. Urinary incontinence may occur. Fecal incontinence is very rare, however. Postictal confusion is also rare, although visual and auditory hallucinations and near-death and out-of-body experiences are sometimes reported.

Although some predisposing factors and provocative stimuli are well established (for example, motionless upright posture, warm ambient temperature, intravascular volume depletion, alcohol ingestion, hypoxemia, anemia, pain, the sight of blood, venipuncture, and intense emotion), the underlying basis for the widely different thresholds for syncope among individuals exposed to the same provocative stimulus is not known. A genetic basis for neurally mediated syncope may exist; several studies have reported an increased incidence of syncope in first-degree relatives of fainters, but no gene or genetic marker has been identified, and environmental, social, and cultural factors have not been excluded by these studies.

TREATMENT

Neurally Mediated Syncope

Reassurance, education, avoidance of provocative stimuli, and plasma volume expansion with fluid and salt are the cornerstones of the management of neurally mediated syncope. Isometric counterpressure maneuvers of the limbs (tensing of the abdominal and leg muscles, handgrip and arm tensing, and leg crossing) may raise blood pressure by increasing central blood volume and cardiac output. Of these, abdominal muscle tensing is the most effective. By maintaining pressure in the autoregulatory zone, these maneuvers, which may be particularly helpful in patients with a long prodrome, avoid or delay the onset of syncope. Randomized controlled trials support this intervention.

Fludrocortisone, vasoconstricting agents, and β -adrenoreceptor antagonists are widely used by experts to treat refractory patients, although there is no consistent evidence from randomized controlled trials for any pharmacotherapy to treat neurally mediated syncope. Because vasodilation, decreased central blood volume, decreased stroke volume and cardiac output are the dominant pathophysiologic syncopal mechanisms in most patients, use of a cardiac pacemaker is rarely beneficial. A systematic review of the literature examining whether cardiac pacing reduces risk of recurrent syncope and relevant clinical outcomes in adults with neurally mediated syncope, concluded that the existing evidence does not support the use of routine cardiac pacing. Possible exceptions are (1) older patients (>40 years), with at least three prior episodes associated with asystole (of at least 3 s associated with syncope or at least 6 s associated with presyncope) documented by an implantable loop recorder; and (2) patients with prominent cardioinhibition due to carotid sinus syndrome. In these patients, dual-chamber pacing may be helpful, although this continues to be an area of uncertainty.

■ ORTHOSTATIC HYPOTENSION

Orthostatic hypotension, defined as a reduction in systolic blood pressure of at least 20 mmHg or diastolic blood pressure of at least 10 mmHg after 3 min of standing or head-up tilt on a tilt table, is a manifestation of sympathetic vasoconstrictor (autonomic) failure (Fig. 21-4). In many (but not all) cases, there is no compensatory

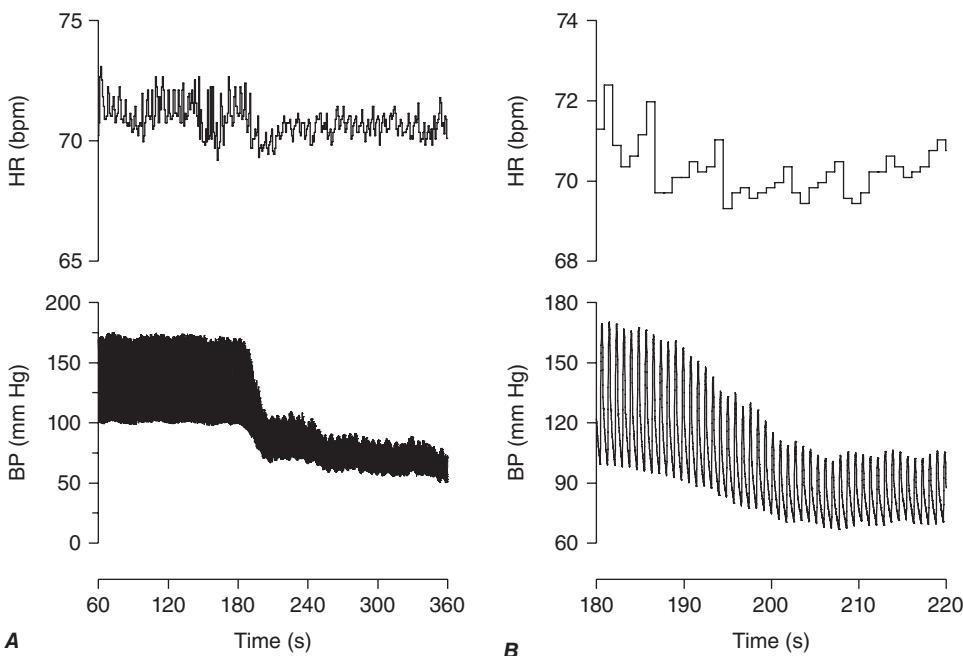


FIGURE 21-4 **A.** The gradual fall in blood pressure without a compensatory heart rate increase that is characteristic of orthostatic hypotension due to autonomic failure. Blood pressure and heart rate are shown >5 min (from 60 to 360 s) of an upright tilt on a tilt table. **B.** The same tracing expanded to show 40 s of the episode (from 180 to 220 s). BP, blood pressure; bpm, beats per minute; HR, heart rate.

increase in heart rate despite hypotension; with partial autonomic failure, heart rate may increase to some degree but is insufficient to maintain cardiac output. A variant of orthostatic hypotension is “delayed” orthostatic hypotension, which occurs beyond 3 min of standing; this may reflect a mild or early form of sympathetic adrenergic dysfunction. In some cases, orthostatic hypotension occurs within 15 s of standing (so-called initial orthostatic hypotension), a finding that may reflect a transient mismatch between cardiac output and peripheral vascular resistance and does not represent autonomic failure.

Characteristic symptoms of orthostatic hypotension include light-headedness, dizziness, and presyncope (near-faintness) occurring in response to sudden postural change. However, symptoms may be absent or nonspecific, such as generalized weakness, fatigue, cognitive slowing, leg buckling, or headache. Visual blurring may occur, likely due to retinal or occipital lobe ischemia. Neck pain, typically in the suboccipital, posterior cervical, and shoulder region (the “coat-hanger headache”), most likely due to neck muscle ischemia, may be the only symptom. Patients may report orthostatic dyspnea (thought to reflect ventilation-perfusion mismatch due to inadequate perfusion of ventilated lung apices) or angina (attributed to impaired myocardial perfusion even with normal coronary arteries). Symptoms may be exacerbated by exertion, prolonged standing, increased ambient temperature, or meals. Syncope is usually preceded by warning symptoms, but may occur suddenly, suggesting the possibility of a seizure or cardiac cause. Some patients have profound decreases in blood pressure, sometimes without symptoms but placing them at risk for falls and injuries if the autoregulatory threshold is crossed with ensuing cerebral hypoperfusion.

Supine hypertension is common in patients with orthostatic hypotension due to autonomic failure, affecting >50% of patients in some series. Orthostatic hypotension may present after initiation of therapy for hypertension, and supine hypertension may follow treatment of orthostatic hypotension. However, in other cases, the association of the two conditions is unrelated to therapy; it may in part be explained by baroreflex dysfunction in the presence of residual sympathetic outflow, particularly in patients with central autonomic degeneration.

Causes of Neurogenic Orthostatic Hypotension Causes of neurogenic orthostatic hypotension include central and peripheral

autonomic nervous system dysfunction ([Chap. 440](#)). Autonomic dysfunction of other organ systems (including the bladder, bowels, sexual organs, and sudomotor system) of varying severity frequently accompanies orthostatic hypotension in these disorders (Table 21-2).

The primary autonomic degenerative disorders are multiple system atrophy (Shy-Drager syndrome; [Chap. 440](#)), Parkinson’s disease ([Chap. 435](#)), dementia with Lewy bodies ([Chap. 434](#)), and pure autonomic failure ([Chap. 440](#)). These are often grouped together as “synucleinopathies” due to the presence of α -synuclein, a protein that aggregates predominantly in the cytoplasm of neurons in the Lewy body disorders (Parkinson’s disease, dementia with Lewy bodies, and pure autonomic failure) and in the glia in multiple system atrophy.

Peripheral autonomic dysfunction may also accompany small-fiber peripheral neuropathies such as those associated with diabetes mellitus, acquired and hereditary amyloidosis, immune-mediated neuropathies, and hereditary sensory and autonomic neuropathies (HSAN; particularly HSAN type III, familial dysautonomia)

([Chaps. 446 and 447](#)). Less frequently, orthostatic hypotension is associated with the peripheral neuropathies that accompany vitamin B_{12} deficiency, neurotoxin exposure, HIV and other infections, and porphyria.

Patients with autonomic failure and the elderly are susceptible to falls in blood pressure associated with meals. The magnitude of the blood pressure fall is exacerbated by large meals, meals high in carbohydrate, and alcohol intake. The mechanism of postprandial syncope is not fully elucidated.

Orthostatic hypotension is often iatrogenic. Drugs from several classes may lower peripheral resistance (e.g., α -adrenoreceptor antagonists used to treat hypertension and prostatic hypertrophy; antihypertensive agents of several classes; nitrates and other vasodilators; tricyclic agents and phenothiazines). Iatrogenic volume depletion due to diuresis and volume depletion due to medical causes (hemorrhage, vomiting, diarrhea, or decreased fluid intake) may also result in decreased effective circulatory volume, orthostatic hypotension, and syncope.

TREATMENT

Orthostatic Hypotension

The first step is to remove reversible causes—usually vasoactive medications ([see Table 440-6](#)). Next, nonpharmacologic interventions should be introduced. These include patient education regarding staged moves from supine to upright; warnings about the hypotensive effects of large meals; instructions about the isometric counterpressure maneuvers that increase intravascular pressure (see above); and raising the head of the bed to reduce supine hypertension and nocturnal diuresis. Intravascular volume should be expanded by increasing dietary fluid and salt. If these nonpharmacologic measures fail, pharmacologic intervention with fludrocortisone acetate and vasoconstricting agents such as midodrine and L-dihydroxyphenylserine should be introduced. Some patients with intractable symptoms require additional therapy with supplementary agents that include pyridostigmine, atomoxetine, yohimbine, octreotide, desmopressin acetate (DDAVP), and erythropoietin ([Chap. 440](#)).

CARDIAC SYNCOPE

Cardiac (or cardiovascular) syncope is caused by arrhythmias and structural heart disease. These may occur in combination because structural disease renders the heart more vulnerable to abnormal electrical activity.

Arrhythmias Bradyarrhythmias that cause syncope include those due to severe sinus node dysfunction (e.g., sinus arrest or sinoatrial block) and atrioventricular (AV) block (e.g., Mobitz type II, high-grade, and complete AV block). The bradyarrhythmias due to sinus node dysfunction are often associated with an atrial tachyarrhythmia, a disorder known as the tachycardia-bradycardia syndrome. A prolonged pause following the termination of a tachycardic episode is a frequent cause of syncope in patients with the tachycardia-bradycardia syndrome. Medications of several classes may also cause bradyarrhythmias of sufficient severity to cause syncope. Syncope due to bradycardia or asystole has been referred to as a Stokes-Adams attack.

Ventricular tachyarrhythmias frequently cause syncope. The likelihood of syncope with ventricular tachycardia is in part dependent on the ventricular rate; rates <200 beats/min are less likely to cause syncope. The compromised hemodynamic function during ventricular tachycardia is caused by ineffective ventricular contraction, reduced diastolic filling due to abbreviated filling periods, loss of AV synchrony, and concurrent myocardial ischemia.

Several disorders associated with cardiac electrophysiologic instability and arrhythmogenesis are due to mutations in ion channel subunit genes. These include the long QT syndrome, Brugada syndrome, and catecholaminergic polymorphic ventricular tachycardia. The long QT syndrome is a genetically heterogeneous disorder associated with prolonged cardiac repolarization and a predisposition to ventricular arrhythmias. Syncope and sudden death in patients with long QT syndrome result from a unique polymorphic ventricular tachycardia called *torsades des pointes* that degenerates into ventricular fibrillation. The long QT syndrome has been linked to genes encoding K⁺ channel α-subunits, K⁺ channel β-subunits, voltage-gated Na⁺ channel, and a scaffolding protein, ankyrin B (ANK2). Brugada syndrome is characterized by idiopathic ventricular fibrillation in association with right ventricular electrocardiogram (ECG) abnormalities without structural heart disease. This disorder is also genetically heterogeneous, although it is most frequently linked to mutations in the Na⁺ channel α-subunit, SCN5A. Catecholaminergic polymorphic tachycardia is an inherited, genetically heterogeneous disorder associated with exercise- or stress-induced ventricular arrhythmias, syncope, or sudden death. Acquired QT interval prolongation, most commonly due to drugs, may also result in ventricular arrhythmias and syncope. **These disorders are discussed in detail in Chap. 255.**

Structural Disease Structural heart disease (e.g., valvular disease, myocardial ischemia, hypertrophic and other cardiomyopathies, cardiac masses such as atrial myxoma, and pericardial effusions) may lead to syncope by compromising cardiac output. Structural disease may also contribute to other pathophysiologic mechanisms of syncope. For example, cardiac structural disease may predispose to arrhythmogenesis; aggressive treatment of cardiac failure with diuretics and/or vasodilators may lead to orthostatic hypotension; and inappropriate reflex vasodilation may occur with structural disorders such as aortic stenosis and hypertrophic cardiomyopathy, possibly provoked by increased ventricular contractility.

TREATMENT

Cardiac Syncope

Treatment of cardiac disease depends on the underlying disorder. Therapies for arrhythmias include cardiac pacing for sinus node disease and AV block, and ablation, antiarrhythmic drugs, and cardioverter-defibrillators for atrial and ventricular tachyarrhythmias. These disorders are best managed by physicians with specialized skills in this area.

APPROACH TO THE PATIENT

Syncope

DIFFERENTIAL DIAGNOSIS

Syncope is easily diagnosed when the characteristic features are present; however, several disorders with transient real or apparent loss of consciousness may create diagnostic confusion.

Generalized and partial seizures may be confused with syncope; however, there are a number of differentiating features. Whereas tonic-clonic movements are the hallmark of a generalized seizure, myoclonic and other movements also may occur in up to 90% of syncopal episodes. Myoclonic jerks associated with syncope may be multifocal or generalized. They are typically arrhythmic and of short duration (<30 s). Mild flexor and extensor posturing also may occur. Partial or partial-complex seizures with secondary generalization are usually preceded by an aura, commonly an unpleasant smell; fear; anxiety; abdominal discomfort; or other visceral sensations. These phenomena should be differentiated from the premonitory features of syncope.

Autonomic manifestations of seizures (autonomic epilepsy) may provide a more difficult diagnostic challenge. Autonomic seizures have cardiovascular, gastrointestinal, pulmonary, urogenital, pupillary, and cutaneous manifestations that are similar to the premonitory features of syncope. Furthermore, the cardiovascular manifestations of autonomic epilepsy include clinically significant tachycardias and bradycardias that may be of sufficient magnitude to cause loss of consciousness. The presence of accompanying non-autonomic auras may help differentiate these episodes from syncope.

Loss of consciousness associated with a seizure usually lasts >5 min and is associated with prolonged postictal drowsiness and disorientation, whereas reorientation occurs almost immediately after a syncopal event. Muscle aches may occur after both syncope and seizures, although they tend to last longer and be more severe following a seizure. Seizures, unlike syncope, are rarely provoked by emotions or pain. Incontinence of urine may occur with both seizures and syncope; however, fecal incontinence occurs very rarely with syncope.

Hypoglycemia may cause transient loss of consciousness, typically in individuals with type 1 or type 2 diabetes (**Chap. 403**) treated with insulin. The clinical features associated with impending or actual hypoglycemia include tremor, palpitations, anxiety, diaphoresis, hunger, and paresthesias. These symptoms are due to autonomic activation to counter the falling blood glucose. Hunger, in particular, is not a typical premonitory feature of syncope. Hypoglycemia also impairs neuronal function, leading to fatigue, weakness, dizziness, and cognitive and behavioral symptoms. Diagnostic difficulties may occur in individuals in strict glycemic control; repeated hypoglycemia impairs the counterregulatory response and leads to a loss of the characteristic warning symptoms that are the hallmark of hypoglycemia.

Patients with cataplexy (**Chap. 31**) experience an abrupt partial or complete loss of muscular tone triggered by strong emotions, typically anger or laughter. Unlike syncope, consciousness is maintained throughout the attacks, which typically last between 30 s and 2 min. There are no premonitory symptoms. Cataplexy occurs in 60%–75% of patients with narcolepsy.

The clinical interview and interrogation of eyewitnesses usually allow differentiation of syncope from falls due to vestibular dysfunction, cerebellar disease, extrapyramidal system dysfunction, and other gait disorders. A diagnosis of syncope can be particularly challenging in patients with dementia who experience repeated falls and are unable to provide a clear history of the episodes. If the fall is accompanied by head trauma, a postconcussive syndrome, amnesia for the precipitating events, and/or a loss or alteration of consciousness, this may also contribute to diagnostic difficulty.

Apparent loss of consciousness can be a manifestation of psychiatric disorders such as generalized anxiety, panic disorders, major

depression, and somatization disorder. These possibilities should be considered in individuals who faint frequently without prodromal symptoms. Such patients are rarely injured despite numerous falls. There are no clinically significant hemodynamic changes concurrent with these episodes. In contrast, transient loss of consciousness due to vasovagal syncope precipitated by fear, stress, anxiety, and emotional distress is accompanied by hypotension, bradycardia, or both.

INITIAL EVALUATION

The goals of the initial evaluation are to determine whether the transient loss of consciousness was due to syncope; to identify the cause; and to assess risk for future episodes and serious harm (Table 21-1). The initial evaluation should include a detailed history, thorough questioning of eyewitnesses, and a complete physical and neurologic examination. Blood pressure and heart rate should be measured in the supine position and after 3 min of standing to determine whether orthostatic hypotension is present. High-risk features on history include: the new onset of chest discomfort, abdominal pain, shortness of breath or headache; syncope during exertion or while supine; sudden onset of palpitations followed by syncope; severe coronary artery or structural heart disease.

High-risk features on examination include an unexplained systolic BP of <90 mmHg; suggestion of gastrointestinal hemorrhage; persistent bradycardia (<40 beats/min); and an undiagnosed systolic murmur.

An ECG should be performed if there is suspicion of syncope due to an arrhythmia or underlying cardiac disease. Relevant electrocardiographic abnormalities include bradyarrhythmias or tachyarrhythmias, AV block, acute myocardial ischemia, old myocardial infarction, long QT_c, and bundle branch block. This initial assessment will lead to the identification of a cause of syncope in ~50% of patients and also allows stratification of patients at risk for cardiac mortality.

Laboratory Tests Baseline laboratory blood tests are rarely helpful in identifying the cause of syncope. Blood tests should be performed when specific disorders, e.g., myocardial infarction, anemia, and secondary autonomic failure, are suspected (Table 21-2).

Autonomic Nervous System Testing (Chap. 440) Autonomic testing, including tilt-table testing, can be performed in specialized centers. Autonomic testing is helpful to uncover objective evidence of autonomic failure and also to demonstrate a predisposition to neurally mediated syncope. Autonomic testing includes assessments of parasympathetic autonomic nervous system function (e.g., heart rate variability to deep respiration and a Valsalva maneuver), sympathetic cholinergic function (e.g., thermoregulatory sweat response and quantitative sudomotor axon reflex test), and sympathetic adrenergic function (e.g., blood pressure response to a Valsalva maneuver and a tilt-table test with beat-to-beat blood pressure measurement). The hemodynamic abnormalities demonstrated on the tilt-table test (Figs. 21-3 and 21-4) may be useful in distinguishing orthostatic hypotension due to autonomic failure from the hypotensive bradycardic response of neurally mediated syncope. Similarly, the tilt-table test may help identify patients with syncope due to immediate or delayed orthostatic hypotension.

Carotid sinus massage should be considered in patients with symptoms suggestive of carotid sinus syncope and in patients >40 years with recurrent syncope of unknown etiology. This test should only be carried out under continuous ECG and blood pressure monitoring and should be avoided in patients with carotid bruits, possible or known plaques, or stenosis.

Cardiac Evaluation ECG monitoring is indicated for patients with a high pretest probability of arrhythmia causing syncope. Patients should be monitored in the hospital if the likelihood of a life-threatening arrhythmia is high, e.g., patients with severe coronary artery or structural heart disease, nonsustained ventricular

tachycardia, supraventricular tachycardia, paroxysmal atrial fibrillation, trifascicular heart block, prolonged QT interval, Brugada syndrome ECG pattern, syncope during exertion, syncope while seated or supine, and family history of sudden cardiac death (Table 21-1). Outpatient Holter monitoring is recommended for patients who experience frequent syncopal episodes (e.g., one or more per week), whereas loop recorders, which continually record and erase cardiac rhythm, are indicated for patients with suspected arrhythmias with low risk of sudden cardiac death. Loop recorders may be external (e.g., for evaluation of episodes that occur at a frequency of >1 per month) or implantable (e.g., if syncope occurs less frequently).

Echocardiography should be performed in patients with a history of cardiac disease or if abnormalities are found on physical examination or the ECG. Echocardiographic diagnoses that may be responsible for syncope include aortic stenosis, hypertrophic cardiomyopathy, cardiac tumors, aortic dissection, and pericardial tamponade. Echocardiography also has a role in risk stratification based on the left ventricular ejection fraction.

Treadmill exercise testing with ECG and blood pressure monitoring should be performed in patients who have experienced syncope during or shortly after exercise. Treadmill testing may help identify exercise-induced arrhythmias (e.g., tachycardia-related AV block) and exercise-induced exaggerated vasodilation.

Electrophysiologic studies are indicated in patients with structural heart disease and ECG abnormalities in whom noninvasive investigations have failed to yield a diagnosis. Electrophysiologic studies have low sensitivity and specificity and should only be performed when a high pretest probability exists. Currently, these tests are rarely performed to evaluate patients with syncope.

Psychiatric Evaluation Screening for psychiatric disorders may be appropriate in patients with recurrent unexplained syncope episodes. Tilt-table testing, with demonstration of symptoms in the absence of hemodynamic change, may be useful in reproducing syncope in patients with suspected psychogenic syncope.

FURTHER READING

- BRIGNOLE M et al: 2018 ESC Guidelines for the diagnosis and management of syncope. Eur Heart J 39:1883, 2018.
- CHESHIRE WP et al: Electrodiagnostic assessment of the autonomic nervous system: a consensus statement endorsed by the American Autonomic Society, American Academy of Neurology, and the International Federation of Clinical Neurophysiology. Clin Neurophysiol 132:666, 2021.
- FREEMAN R et al: Consensus statement on the definition of orthostatic hypotension, neurally mediated syncope and the postural tachycardia syndrome. Auton Neurosci 161:46, 2011.
- FREEMAN R et al: Orthostatic Hypotension: JACC State-of-the-Art Review. J Am Coll Cardiol 72:1294, 2018.
- GIBBONS CH et al: The recommendations of a consensus panel for the screening, diagnosis, and treatment of neurogenic orthostatic hypotension and associated supine hypertension. J Neurol 264:1567, 2017.
- SHELDON RS, RAJ SR: Pacing and vasovagal syncope: back to our physiologic roots. Clin Auton Res 27:213, 2017.
- SHEN WK et al: 2017 ACC/AHA/HRS Guideline for the Evaluation and Management of Patients With Syncope: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. Circulation 136:e60, 2017.
- VAROSY PD et al: Pacing as a treatment for reflex-mediated (vasovagal, situational, or carotid sinus hypersensitivity) syncope: a systematic review for the 2017 ACC/AHA/HRS guideline for the evaluation and management of patients with syncope: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. J Am Coll Cardiol 70:664, 2017.

22**Dizziness and Vertigo**

Mark F. Walker, Robert B. Daroff



Dizziness is an imprecise symptom used to describe a variety of common sensations that include vertigo, light-headedness, faintness, and imbalance. *Vertigo* refers to a sense of spinning or other motion that may be physiological, occurring during or after a sustained head rotation, or pathological, due to vestibular dysfunction. The term *light-headedness* is classically applied to presyncopal sensations resulting from brain hypoperfusion but as used by patients has little specificity, as it may also refer to other symptoms such as disequilibrium and imbalance. A challenge to diagnosis is that patients often have difficulty distinguishing among these various symptoms, and the words they choose do not reliably indicate the underlying etiology.

There are many causes of dizziness. Vestibular dizziness (vertigo or imbalance) may be due to peripheral disorders that affect the labyrinths or vestibular nerves, or it may result from disruption of central vestibular pathways. It may be paroxysmal or due to a fixed unilateral or bilateral vestibular deficit. Acute unilateral lesions cause vertigo due to a sudden imbalance in vestibular inputs from the two labyrinths. Bilateral lesions cause imbalance and instability of vision when the head moves (*oscillopsia*) due to loss of normal vestibular reflexes.

Presyncopal dizziness occurs when cardiac dysrhythmia, orthostatic hypotension, medication effects, or another cause leads to brain hypoperfusion. Such presyncopal sensations vary in duration; they may increase in severity until loss of consciousness occurs, or they may resolve before loss of consciousness if the cerebral ischemia is corrected. Faintness and syncope, which are discussed in detail in **Chap. 21**, should always be considered when one is evaluating patients with brief episodes of dizziness or dizziness that occurs with upright posture. Other causes of dizziness include nonvestibular imbalance, gait disorders (e.g., loss of proprioception from sensory neuropathy, parkinsonism), and anxiety.

When evaluating patients with dizziness, questions to consider include the following: (1) Is it dangerous (e.g., arrhythmia, transient ischemic attack/stroke)? (2) Is it vestibular? (3) If vestibular, is it peripheral or central? A careful history and examination often provide sufficient information to answer these questions and determine whether additional studies or referral to a specialist is necessary.

APPROACH TO THE PATIENT

Dizziness

HISTORY

When a patient presents with dizziness, the first step is to delineate more precisely the nature of the symptom. In the case of vestibular disorders, the physical symptoms depend on whether the lesion is unilateral or bilateral, and whether it is acute or chronic. Vertigo, an illusion of self or environmental motion, implies an acute asymmetry of vestibular inputs from the two labyrinths or in their central pathways. Symmetric bilateral vestibular hypofunction causes imbalance but no vertigo. Because of the ambiguity in patients' descriptions of their symptoms, diagnosis based simply on symptom characteristics is typically unreliable. Thus the history should focus closely on other features, including whether this is the first attack, the duration of this and any prior episodes, provoking factors, and accompanying symptoms.

Dizziness can be divided into episodes that last for seconds, minutes, hours, or days. Common causes of brief dizziness (seconds) include benign paroxysmal positional vertigo (BPPV) and orthostatic hypotension, both of which typically are provoked by changes in head and/or body position relative to gravity. Attacks of vestibular migraine and Ménière's disease often last hours. When episodes are of intermediate duration (minutes), transient ischemic

attacks of the posterior circulation should be considered, although migraine and other causes are also possible.

Symptoms that accompany vertigo may be helpful in distinguishing peripheral vestibular lesions from central causes. Unilateral hearing loss and other acute aural symptoms (ear pain, pressure, fullness, new tinnitus) typically point to a peripheral cause. Because the auditory pathways quickly become bilateral upon entering the brainstem, central lesions are unlikely to cause unilateral hearing loss unless the lesion lies near the root entry zone of the auditory nerve. Symptoms such as double vision, numbness, and limb ataxia suggest a brainstem or cerebellar lesion.

EXAMINATION

Because dizziness and imbalance can be a manifestation of a variety of neurologic disorders, the neurologic examination is important in the evaluation of these patients. Focus should be given to assessment of eye movements, vestibular function, and hearing. The range of eye movements and whether they are equal in each eye should be observed. Peripheral eye movement disorders (e.g., cranial neuropathies, eye muscle weakness) are usually disconjugate (different in the two eyes). One should check pursuit (the ability to follow a smoothly moving target) and saccades (the ability to look back and forth accurately between two targets). Poor pursuit or inaccurate (dysmetric) saccades usually indicate central pathology, often involving the cerebellum. Alignment of the two eyes can be checked with a cover test: while the patient is looking at a target, alternately cover the eyes and observe for corrective saccades. A vertical misalignment may indicate a brainstem or cerebellar lesion. Finally, one should look for spontaneous nystagmus, an involuntary back-and-forth movement of the eyes. Nystagmus is most often of the jerk type, in which a slow drift (slow phase) in one direction alternates with a rapid saccadic movement (quick phase or fast phase) in the opposite direction that resets the position of the eyes in the orbits. Except in the case of acute vestibulopathy (e.g., vestibular neuritis), if primary position nystagmus is easily seen in the light, it is probably due to a central cause. Two forms of nystagmus that are characteristic of lesions of the cerebellar pathways are vertical nystagmus with downward fast phases (downbeat nystagmus) and horizontal nystagmus that changes direction with gaze (gaze-evoked nystagmus). By contrast, peripheral lesions typically cause unidirectional horizontal nystagmus. Use of Frenzel eyeglasses (self-illuminated goggles with convex lenses that blur the patient's vision but allow the examiner to see the eyes greatly magnified) or infrared video goggles can aid in the detection of peripheral vestibular nystagmus, because they reduce the patient's ability to use visual fixation to suppress nystagmus. **Table 22-1** outlines key findings that help distinguish peripheral from central causes of vertigo.

The most useful bedside test of peripheral vestibular function is the head impulse test, in which the vestibulo-ocular reflex (VOR) is assessed with small-amplitude (~20 degrees) rapid head rotations. While the patient fixates on a target, the head is rotated quickly to the left or right. If the VOR is deficient, the rotation is followed by a catch-up saccade in the opposite direction (e.g., a leftward saccade

TABLE 22-1 Features of Peripheral and Central Vertigo

- Nystagmus from an acute peripheral lesion is unidirectional, with fast phases beating away from the ear with the lesion. Nystagmus that changes direction with gaze is due to a central lesion.
- Transient mixed vertical-torsional nystagmus occurs in benign paroxysmal positional vertigo (BPPV), but pure vertical or pure torsional nystagmus is a central sign.
- Nystagmus from a peripheral lesion may be inhibited by visual fixation, whereas central nystagmus is not suppressed.
- Absence of a head impulse sign in a patient with acute prolonged vertigo should suggest a central cause.
- Unilateral hearing loss suggests peripheral vertigo. Findings such as diplopia, dysarthria, and limb ataxia suggest a central disorder.

after a rightward rotation). The head impulse test can identify both unilateral (catch-up saccades after rotations toward the weak side) and bilateral (catch-up saccades after rotations in both directions) vestibular hypofunction.

All patients with episodic dizziness, especially if provoked by positional change, should be tested with the Dix-Hallpike maneuver. The patient begins in a sitting position with the head turned 45 degrees; holding the back of the head, the examiner then lowers the patient into a supine position with the head extended backward by about 20 degrees while watching the eyes. Posterior canal BPPV can be diagnosed confidently if transient upbeat-torsional nystagmus is seen. If no nystagmus is observed after 15–20 s, the patient is raised to the sitting position, and the procedure is repeated with the head turned to the other side. Again, Frenzel goggles may improve the sensitivity of the test.

Dynamic visual acuity is a functional test that can be useful in assessing vestibular function. Visual acuity is measured with the head still and when the head is rotated back and forth by the examiner (about 1–2 Hz). A drop in visual acuity during head motion of more than one line on a near card or Snellen chart is abnormal and indicates vestibular dysfunction.

ANCILLARY TESTING

The choice of ancillary tests should be guided by the history and examination findings. Audiometry should be performed whenever a vestibular disorder is suspected. Unilateral sensorineural hearing loss supports a peripheral disorder (e.g., vestibular schwannoma). Predominantly low-frequency hearing loss is characteristic of Ménière's disease. Videonystagmography includes recordings of spontaneous nystagmus (if present) and measurement of positional nystagmus. Caloric testing compares the responses of the two horizontal semicircular canals, while video head-impulse testing measures the integrity of each of the six semicircular canals. Vestibular evoked potentials assess otolith reflexes. The test battery often includes recording of saccades and pursuit to evaluate central ocular motor function. Neuroimaging is important if a central vestibular disorder is suspected. In addition, patients with unexplained unilateral hearing loss or vestibular hypofunction should undergo MRI of the internal auditory canals, including administration of gadolinium, to rule out a schwannoma.

DIFFERENTIAL DIAGNOSIS AND TREATMENT

Treatment of vestibular symptoms should be driven by the underlying diagnosis. Simply treating dizziness with vestibular suppressant medications is often not helpful and may make the symptoms worse and prolong recovery. The diagnostic and specific treatment approaches for the most commonly encountered vestibular disorders are discussed below.

ACUTE PROLONGED VERTIGO (VESTIBULAR NEURITIS)

An acute unilateral vestibular lesion causes constant vertigo, nausea, vomiting, oscillopsia (motion of the visual scene), and imbalance. These symptoms are due to a sudden asymmetry of inputs from the two labyrinths or in their central connections, simulating a continuous rotation of the head. Unlike BPPV, continuous vertigo persists even when the head remains still.

When a patient presents with an acute vestibular syndrome, the most important question is whether the lesion is central (e.g., a cerebellar or brainstem infarct or hemorrhage), which may be life-threatening, or peripheral, affecting the vestibular nerve or labyrinth (vestibular neuritis). Attention should be given to any symptoms or signs that point to central dysfunction (diplopia, weakness or numbness, dysarthria). The pattern of spontaneous nystagmus, if present, may be helpful (Table 22-1). If the head impulse test is normal, an acute peripheral vestibular lesion is unlikely. A central lesion cannot always be excluded with certainty based on symptoms and examination alone; thus older patients with vascular risk factors who present with an acute vestibular

syndrome should be evaluated for the possibility of stroke even when there are no specific findings that indicate a central lesion.

Most patients with vestibular neuritis recover spontaneously, although chronic dizziness, motion sensitivity, and disequilibrium may persist. The role of early glucocorticoid therapy is uncertain, as studies have yielded disparate results. Antiviral medications are of no proven benefit and are not typically given unless there is evidence to suggest herpes zoster oticus (Ramsay Hunt syndrome). Vestibular suppressant medications may reduce acute symptoms but should be avoided after the first several days because they may impede central compensation and recovery. Patients should be encouraged to resume a normal level of activity as soon as possible, and directed vestibular rehabilitation therapy may accelerate improvement.

BENIGN PAROXYSMAL POSITIONAL VERTIGO

BPPV is a common cause of recurrent vertigo. Episodes are brief (<1 min and typically 15–20 s) and are always provoked by changes in head position relative to gravity, such as lying down, rising from a supine position, and extending the head to look upward. Rolling over in bed is a common trigger that may help to distinguish BPPV from orthostatic hypotension. The attacks are caused by free-floating otoconia (calcium carbonate crystals) that have been dislodged from the utricular macula and have moved into one of the semicircular canals, usually the posterior canal. When head position changes, gravity causes the otoconia to move within the canal, producing vertigo and nystagmus. With posterior canal BPPV, the nystagmus beats upward and torsionally (the upper poles of the eyes beat toward the affected lower ear). Less commonly, the otoconia enter the horizontal canal, resulting in a horizontal nystagmus when the patient is lying with either ear down. Superior (also called anterior) canal involvement is rare. BPPV is treated with repositioning maneuvers that use gravity to remove the otoconia from the semicircular canal. For posterior canal BPPV, the Epley maneuver (Fig. 22-1) is the most commonly used procedure. For more refractory cases of BPPV, patients can be taught a variant of this maneuver that they can perform alone at home. A demonstration of the Epley maneuver is available online (<http://www.dizziness-and-balance.com/disorders/bppv/bppv.html>).

VESTIBULAR MIGRAINE

Vestibular migraine is a common yet underdiagnosed cause of episodic vertigo. Vertigo sometimes precedes a typical migraine headache but more often occurs without headache or with only a mild headache. Some patients who have had frequent migraine headaches in the past present later in life with vestibular migraine as the predominant problem. In vestibular migraine, the duration of vertigo may be from minutes to hours, and some migraineurs also experience more prolonged periods of disequilibrium (lasting days to weeks). Motion sensitivity and sensitivity to visual motion (e.g., movies) are common. Even in the absence of headache, other migraine features may be present, such as photophobia, phonophobia, or a visual aura. Although data from controlled studies are generally lacking, vestibular migraine typically is treated with medications that are used for prophylaxis of migraine headaches (Chap. 430). Antiemetics may be helpful to relieve symptoms at the time of an attack.

MÉNIÈRE'S DISEASE

Attacks of Ménière's disease consist of vertigo and hearing loss, as well as pain, pressure, and/or fullness in the affected ear. Low-frequency hearing loss and aural symptoms are key features that distinguish Ménière's disease from other peripheral vestibulopathies and from vestibular migraine. Audiometry at the time of an attack shows a characteristic asymmetric low-frequency hearing loss; hearing commonly improves between attacks, although permanent hearing loss may eventually occur. Ménière's disease is associated with excess endolymph fluid in the inner ear; hence the term *endolymphatic hydrops*. The exact pathophysiological mechanism, however, remains unclear. Patients suspected of having Ménière's disease should be referred to an otolaryngologist for further evaluation. Diuretics and sodium restriction are typically the initial treatments. If attacks persist, injections of

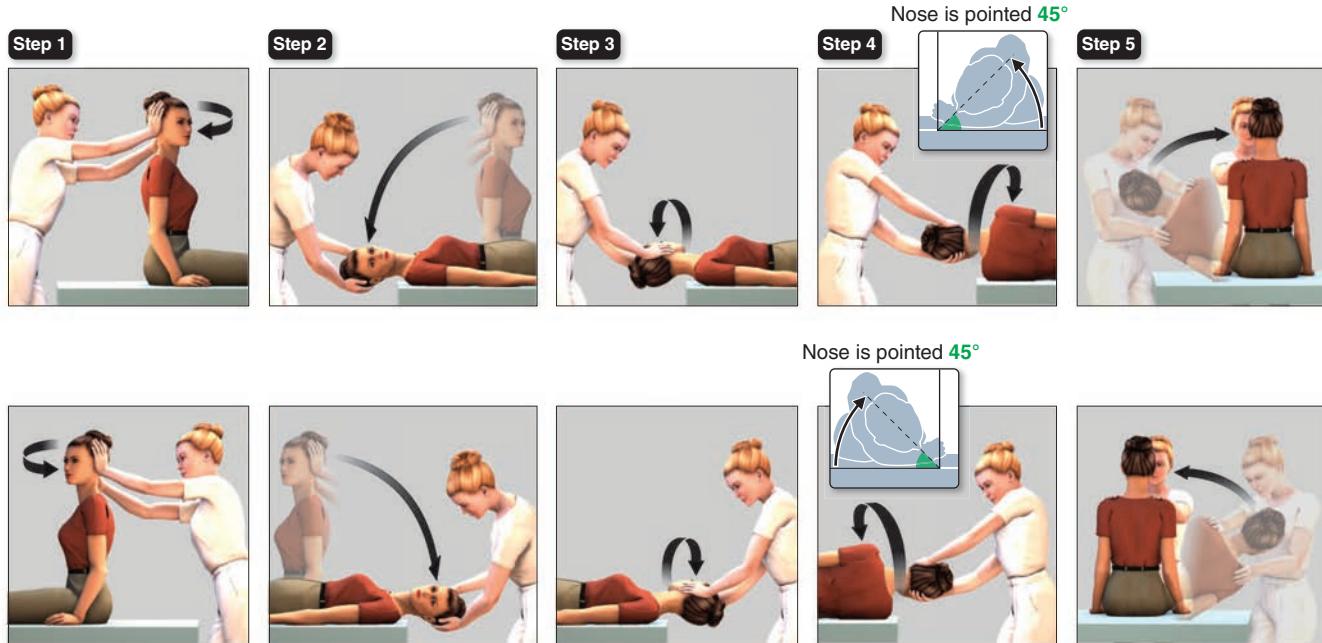


FIGURE 22-1 Modified Epley maneuver for treatment of benign paroxysmal positional vertigo of the right (top panels) and left (bottom panels) posterior semicircular canals. **Step 1.** With the patient seated, turn the head 45 degrees toward the affected ear. **Step 2.** Keeping the head turned, lower the patient to the head-hanging position and hold for at least 30 s and until nystagmus disappears. **Step 3.** Without lifting the head, turn it 90 degrees toward the other side. Hold for another 30 s. **Step 4.** Rotate the patient onto her side while turning the head another 90 degrees, so that the nose is pointed down 45 degrees. Hold again for 30 s. **Step 5.** Have the patient sit up on the side of the table. After a brief rest, the maneuver should be repeated to confirm successful treatment. (Reproduced with permission from Chicago dizziness and Hearing (CDH). Figure adapted from <http://www.dizziness-and-balance.com/disorders/bppv/movies/Epley-480x640.avi>)

glucocorticoids or gentamicin into the middle ear may be considered. Nonablative surgical options include decompression and shunting of the endolymphatic sac. Full ablative procedures (vestibular nerve section, labyrinthectomy) are seldom required.

■ VESTIBULAR SCHWANNOMA

Vestibular schwannomas (sometimes termed *acoustic neuromas*) and other tumors at the cerebellopontine angle cause slowly progressive unilateral sensorineural hearing loss and vestibular hypofunction. These patients typically do not have vertigo, because the gradual vestibular deficit is compensated centrally as it develops. The diagnosis often is not made until there is sufficient hearing loss to be noticed. The vestibular examination will show a deficient response to the head impulse test when the head is rotated toward the affected side, but nystagmus will not be prominent. As noted above, patients with unexplained unilateral sensorineural hearing loss or vestibular hypofunction require MRI of the internal auditory canals to look for a schwannoma.

■ BILATERAL VESTIBULAR HYPOFUNCTION

Patients with bilateral loss of vestibular function also typically do not have vertigo, because vestibular function is lost on both sides simultaneously, and there is no asymmetry of vestibular input. Symptoms include loss of balance, particularly in the dark, where vestibular input is most critical, and oscillopsia during head movement, such as while walking or riding in a car. Bilateral vestibular hypofunction may be (1) idiopathic and progressive, (2) part of a neurodegenerative disorder, or (3) iatrogenic due to medication ototoxicity (most commonly gentamicin or other aminoglycoside antibiotics). Other causes include bilateral vestibular schwannomas (neurofibromatosis type 2), autoimmune disease, superficial siderosis, and meningeal-based infection or tumor. It also may occur in patients with peripheral polyneuropathy; in these patients, both vestibular loss and impaired proprioception may contribute to poor balance. Finally, unilateral processes such as vestibular neuritis and Ménière's disease may involve both ears sequentially, resulting in bilateral vestibulopathy.

Examination findings include diminished *dynamic visual acuity* (see above) due to loss of stable vision when the head is moving, abnormal head impulse responses in both directions, and a Romberg

sign. Responses to caloric testing are reduced. Patients with bilateral vestibular hypofunction should be referred for vestibular rehabilitation therapy. Vestibular suppressant medications should not be used, as they will increase the imbalance. Evaluation by a neurologist is important not only to confirm the diagnosis but also to consider any other associated neurologic abnormalities that may clarify the etiology.

■ CENTRAL VESTIBULAR DISORDERS

Central lesions causing vertigo typically involve vestibular pathways in the brainstem and/or cerebellum. They may be due to discrete lesions, such as from ischemic or hemorrhagic stroke (Chaps. 426–428), demyelination (Chap. 444), or tumors (Chap. 90), or they may be due to neurodegenerative conditions that include the vestibulocerebellum (Chaps. 431–434). Subacute cerebellar degeneration may be due to immune, including paraneoplastic, processes (Chaps. 94 and 439). Table 22-1 outlines important features of the history and examination that help to identify central vestibular disorders. Acute central vertigo is a medical emergency, due to the possibility of life-threatening stroke or hemorrhage. All patients with suspected central vestibular disorders should undergo brain MRI, and the patient should be referred for full neurologic evaluation.

■ PSYCHOSOMATIC AND FUNCTIONAL DIZZINESS

Psychological factors play an important role in chronic dizziness. First, dizziness may be a somatic manifestation of a psychiatric condition such as major depression, anxiety, or panic disorder (Chap. 452). Second, patients may develop anxiety and autonomic symptoms as a consequence or comorbidity of an independent vestibular disorder. One particular form of this has been termed variously *phobic postural vertigo*, *psychophysiological vertigo*, or *chronic subjective dizziness*, but is now referred to as *persistent postural-perceptual dizziness (PPPD)*. These patients have a chronic feeling (3 months or longer) of fluctuating dizziness and disequilibrium that is present at rest but worse while standing. There is an increased sensitivity to self-motion and visual motion (e.g., watching movies), and a particular intensification of symptoms when moving through complex visual environments such as supermarkets. Although there may be a past history of an acute vestibular disorder (e.g., vestibular neuritis), the neuro-otologic examination

TABLE 22-2 Treatment of Vertigo

AGENT ^a	DOSE ^b
Antihistamines	
Meclizine	25–50 mg 3 times daily
Dimenhydrinate	50 mg 1–2 times daily
Promethazine	25 mg 2–3 times daily (also can be given rectally and IM)
Benzodiazepines	
Diazepam	2.5 mg 1–3 times daily
Clonazepam	0.25 mg 1–3 times daily
Anticholinergic	
Scopolamine transdermal ^c	Patch
Physical therapy	
Repositioning maneuvers ^d	
Vestibular rehabilitation	
Other	
Diuretics and/or low-sodium (1000 mg/d) diet ^e	
Antimigrainous drugs ^f	
Selective serotonin reuptake inhibitors ^g	

^aAll listed drugs are approved by the US Food and Drug Administration, but most are not approved for the treatment of vertigo. ^bUsual oral (unless otherwise stated) starting dose in adults; a higher maintenance dose can be reached by a gradual increase. ^cFor motion sickness only. ^dFor benign paroxysmal positional vertigo. ^eFor Ménière's disease. ^fFor vestibular migraine. ^gFor persistent postural-perceptual vertigo and anxiety.

and vestibular testing are normal or indicative of a compensated vestibular deficit, indicating that the ongoing subjective dizziness cannot be explained by a primary vestibular pathology. Anxiety disorders are particularly common in patients with chronic dizziness; when present, they contribute substantially to the morbidity. Treatment approaches for PPPD include pharmacological therapy with selective serotonin reuptake inhibitors (SSRIs), cognitive-behavioral psychotherapy, and vestibular suppressant medications generally should be avoided.

TREATMENT

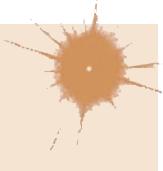
Vertigo

Table 22-2 provides a list of commonly used medications for suppression of vertigo. As noted, these medications should be reserved for short-term control of active vertigo, such as during the first few days of acute vestibular neuritis, or for acute attacks of Ménière's disease. They are less helpful for chronic dizziness and, as previously stated, may hinder central compensation. An exception is that benzodiazepines may attenuate psychosomatic dizziness and the associated anxiety, although SSRIs are generally preferable in such patients.

Vestibular rehabilitation therapy promotes central adaptation processes that compensate for vestibular loss and also may help habituate motion sensitivity and other symptoms of psychosomatic dizziness. The general approach is to use a graded series of exercises that progressively challenge gaze stabilization and balance.

FURTHER READING

- ALTISSIMI G et al: Drugs inducing hearing loss, tinnitus, dizziness and vertigo: An updated guide. *Eur Rev Med Pharmacol Sci* 24:7946, 2020.
- HUANG TC et al: Vestibular migraine: An update on current understanding and future directions. *Cephalgia* 40:107, 2020.
- KIM JS, ZEE DS: Benign paroxysmal positional vertigo. *N Engl J Med* 370:1138, 2014.
- POPKIROV S et al: Persistent postural-perceptual dizziness (PPPD): a common, characteristic and treatable cause of chronic dizziness. *Pract Neurol* 18:5, 2018.



Fatigue is one of the most common symptoms in clinical medicine. It is a prominent manifestation of a number of systemic, neurologic, and psychiatric syndromes, although a precise cause will not be identified in a substantial minority of patients. Fatigue refers to the subjective experience of physical and mental weariness, sluggishness, low energy, and exhaustion. In the context of clinical medicine, fatigue is most practically defined as difficulty initiating or maintaining voluntary mental or physical activity. Nearly everyone who has ever been ill with a self-limited infection has experienced this near-universal symptom, and fatigue is usually brought to medical attention only when it is either of unclear cause, fails to remit, or the severity is out of proportion with what would be expected for the associated trigger.

Fatigue should be distinguished from *muscle weakness*, a reduction of neuromuscular power (**Chap. 24**); most patients complaining of fatigue are not truly weak when direct muscle power is tested. Fatigue is also distinct from *somnolence*, which refers to sleepiness in the context of disturbed sleep-wake physiology (**Chap. 31**), and from *dyspnea on exertion*, although patients may use the word fatigue to describe any of these symptoms. The task facing clinicians when a patient presents with fatigue is to identify the underlying cause and develop a therapeutic alliance, the goal of which is to spare patients expensive and fruitless diagnostic workups and steer them toward effective therapy.

■ EPIDEMIOLOGY AND GLOBAL CONSIDERATIONS

Variability in the definitions of fatigue and the survey instruments used in different studies makes it difficult to arrive at precise figures about the global burden of fatigue. The point prevalence of fatigue was 6.7% and the lifetime prevalence was 25% in a large National Institute of Mental Health survey of the U.S. general population. In primary care clinics in Europe and the United States, between 10 and 25% of patients surveyed endorsed symptoms of prolonged (present for >1 month) or chronic (present for >6 months) fatigue, but in only a minority was fatigue the primary reason for seeking medical attention. In a community survey of women in India, 12% reported chronic fatigue. By contrast, the prevalence of chronic fatigue syndrome (**Chap. 450**), as defined by the U.S. Centers for Disease Control and Prevention, is low.

■ DIFFERENTIAL DIAGNOSIS

Psychiatric Disease Fatigue is a common somatic manifestation of many major psychiatric syndromes, including depression, anxiety, and somatoform disorders (**Chap. 452**). Psychiatric symptoms are reported in more than three-quarters of patients with unexplained chronic fatigue. Even in patients with systemic or neurologic disorders in which fatigue is independently recognized as a symptom, comorbid psychiatric disease may still be an important contributor.

Neurologic Disease Patients complaining of fatigue often say they feel weak, but upon careful examination, objective muscle weakness is rarely discernible. If found, muscle weakness must then be localized to the central nervous system, peripheral nervous system, neuromuscular junction, or muscle, and appropriate follow-up studies obtained (**Chap. 24**). *Fatigability* of muscle power is a cardinal manifestation of some neuromuscular disorders such as myasthenia gravis and is distinguished from *fatigue* by finding clinically evident diminution of the amount of force that a muscle generates upon repeated contraction (**Chap. 448**). Fatigue is one of the most common and bothersome symptoms reported in multiple sclerosis (MS) (**Chap. 444**), affecting nearly 90% of patients; fatigue in MS can persist between MS attacks and does not necessarily correlate with magnetic resonance imaging (MRI) disease activity. Fatigue is also increasingly identified as a troublesome feature of many neurodegenerative diseases, including Parkinson's disease (**Chap. 435**), amyotrophic lateral sclerosis

([Chap. 437](#)), and central nervous system dysautonomias ([Chap. 440](#)). Fatigue after stroke ([Chap. 426](#)) is a well-described but poorly understood entity with a widely varying prevalence. Episodic fatigue can be a premonitory symptom of migraine ([Chap. 430](#)). Fatigue is also a frequent consequence of traumatic brain injury ([Chap. 443](#)), often occurring in association with depression and sleep disorders.

Sleep Disorders Obstructive sleep apnea is an important cause of excessive daytime sleepiness in association with fatigue and should be investigated using overnight polysomnography, particularly in those with prominent snoring, obesity, or other predictors of obstructive sleep apnea ([Chap. 297](#)). Whether the cumulative sleep deprivation that is common in modern society contributes to clinically apparent fatigue is not known ([Chap. 31](#)).

Endocrine Disorders Fatigue, sometimes in association with true muscle weakness, can be a heralding symptom of hypothyroidism ([Chap. 383](#)), particularly in the context of hair loss, dry skin, cold intolerance, constipation, and weight gain. Fatigue associated with heat intolerance, sweating, and palpitations is typical of hyperthyroidism ([Chap. 384](#)). Adrenal insufficiency ([Chap. 386](#)) can also manifest with unexplained fatigue as a primary or prominent symptom, often with anorexia, weight loss, nausea, myalgias, and arthralgias; hyponatremia, hyperkalemia, and hyperpigmentation may be present at time of diagnosis. Mild hypercalcemia can cause fatigue, which may be relatively vague, whereas severe hypercalcemia can lead to lethargy, stupor, and coma ([Chap. 410](#)). Both hypoglycemia and hyperglycemia can cause lethargy, often in association with confusion; diabetes mellitus, and in particular type 1 diabetes, is also associated with fatigue independent of glucose levels ([Chap. 403](#)). Fatigue may also accompany Cushing's disease, hypoaldosteronism, and hypogonadism. Low vitamin D status has also been associated with fatigue.

Liver and Kidney Disease Both chronic liver failure and chronic kidney disease can cause fatigue. Over 80% of hemodialysis patients complain of fatigue, which makes it one of the most common symptoms reported by patients in chronic kidney disease ([Chap. 311](#)).

Obesity Obesity ([Chap. 401](#)) is associated with fatigue and sleepiness independent of the presence of obstructive sleep apnea. Obese patients undergoing bariatric surgery experience improvement in daytime sleepiness sooner than would be expected if the improvement were solely the result of weight loss and resolution of sleep apnea. A number of other factors common in obese patients are likely contributors as well, including physical inactivity, diabetes, and depression.

Physical Inactivity Physical inactivity is associated with fatigue, and increasing physical activity can improve fatigue in some patients.

Malnutrition Although fatigue can be a presenting feature of malnutrition ([Chap. 334](#)), nutritional status may also be an important comorbidity and contributor to fatigue in other chronic illnesses, including cancer-associated fatigue.

Infection Both acute and chronic infections commonly lead to fatigue as part of the broader infectious syndrome. Evaluation for undiagnosed infection as the cause of unexplained fatigue, and particularly prolonged or chronic fatigue, should be guided by the history, physical examination, and infectious risk factors, with particular attention to risk for tuberculosis, HIV, chronic hepatitis, and endocarditis. Infectious mononucleosis may cause prolonged fatigue that persists for weeks to months following the acute illness, but infection with the Epstein-Barr virus is only very rarely the cause of unexplained chronic fatigue. Postinfectious fatigue may also occur following a variety of acute infections. For example, a substantial minority of patients who have recovered from SARS-CoV-1, SARS-CoV-2, and Ebola virus complain of persistent fatigue.

Drugs Many medications, drugs, drug withdrawal, and chronic alcohol use can all lead to fatigue. Medications that are more likely to

be causative include antidepressants, antipsychotics, anxiolytics, opiates, antispasticity agents, antiseizure agents, and beta blockers.

Cardiovascular and Pulmonary Disorders Fatigue is one of the most taxing symptoms reported by patients with congestive heart failure and chronic obstructive pulmonary disease and negatively affects quality of life. In a population-based cohort study in Norfolk, United Kingdom, fatigue was associated with an increased hazard of all-cause mortality in the general population, but particularly for deaths related to cardiovascular disease.

Malignancy Fatigue, particularly in association with unexplained weight loss, can be a sign of occult malignancy, but cancer is rarely identified in patients with unexplained chronic fatigue in the absence of other telltale signs or symptoms. Cancer-related fatigue is experienced by 40% of patients at the time of diagnosis and by >80% at some time in the disease course.

Hematologic Disorders Chronic or progressive anemia may present with fatigue, sometimes in association with exertional tachycardia and breathlessness. Anemia may also contribute to fatigue in chronic illness. Low serum ferritin in the absence of anemia may also cause fatigue that is reversible with iron replacement.

Immune-Mediated Disorders Fatigue is a prominent complaint in many chronic inflammatory disorders, including systemic lupus erythematosus, polymyalgia rheumatica, rheumatoid arthritis, inflammatory bowel disease, antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis, sarcoidosis, and Sjögren's syndrome, but is not usually an isolated symptom. Fatigue is also associated with primary immunodeficiency diseases.

Pregnancy Fatigue is very commonly reported by women during all stages of pregnancy and postpartum.

Disorders of Unclear Cause Myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) ([Chap. 450](#)) and fibromyalgia ([Chap. 373](#)) incorporate chronic fatigue as part of the syndromic definition when fatigue is present in association with other criteria, as discussed in the respective chapters. Chronic multisymptom illness, also known as Gulf-War syndrome, is another symptom complex with prominent fatigue; it is most commonly, although not exclusively, observed in veterans of the 1991 Gulf War conflict ([Chap. 57](#)). Idiopathic chronic fatigue is used to describe the syndrome of unexplained chronic fatigue in the absence of enough additional clinical features to meet the diagnostic criteria for ME/CFS.

APPROACH TO THE PATIENT

Fatigue

A detailed history focusing on the quality, pattern, time course, associated symptoms, and alleviating factors of fatigue is necessary to define the syndrome and help direct further evaluation and treatment. It is important to determine if fatigue is the appropriate designation, whether symptoms are acute or chronic, and if the impairment is primarily mental, physical, or a combination of the two. The review of systems should attempt to distinguish fatigue from excessive sleepiness, dyspnea on exertion, exercise intolerance, and muscle weakness. The presence of fever, chills, night sweats, or weight loss should raise suspicion for an occult infection or malignancy. A careful review of prescription, over-the-counter, herbal, and recreational drug and alcohol use is required. Circumstances surrounding the onset of symptoms and potential triggers should be investigated. The social history is important, with attention paid to life stressors and adverse experiences, workhours, the social support network, and domestic affairs including a screen for intimate partner violence. Sleep habits and sleep hygiene should be questioned. The impact of fatigue on daily functioning is important to understand the patient's experience and gauge recovery and the success of treatment.

The physical examination of patients with fatigue is guided by the history and differential diagnosis. A detailed mental status examination should be performed with particular attention to symptoms of depression and anxiety. A formal neurologic examination is required to determine whether objective muscle weakness is present. This is usually a straightforward exercise, although occasionally patients with fatigue have difficulty sustaining effort against resistance and sometimes report that generating full power requires substantial mental effort. On confrontational testing, full power may be generated for only a brief period before the patient suddenly gives way to the examiner. This type of weakness is often referred to as *breakaway weakness* and may or may not be associated with pain. This is contrasted with weakness due to lesions in the motor tracts or lower motor unit, in which the patient's resistance can be overcome in a smooth and steady fashion and full power can never be generated. Occasionally, a patient may demonstrate fatigable weakness, in which power is full when first tested but becomes weak upon repeat evaluation without interval rest. Fatigable weakness, which usually indicates a problem of neuromuscular transmission, never has the sudden breakaway quality that one occasionally observes in patients with fatigue. If the presence or absence of muscle weakness cannot be determined with the physical examination, electromyography with nerve conduction studies can be a helpful ancillary test.

The general physical examination should screen for signs of cardiopulmonary disease, malignancy, lymphadenopathy, organomegaly, infection, liver failure, kidney disease, malnutrition, endocrine abnormalities, and connective tissue disease. In patients with associated widespread musculoskeletal pain, assessment of tender points may help to reveal fibromyalgia. Although the diagnostic yield of the general physical examination may be relatively low in the context of evaluation of unexplained chronic fatigue, elucidating the cause of only 2% of cases in one prospective analysis, the yield of a detailed neuropsychiatric and mental status evaluation is likely to be much higher, revealing a potential explanation for fatigue in up to 75–80% of patients in some series. Furthermore, a complete physical examination demonstrates a serious and systematic approach to the patient's complaint and helps build trust and a therapeutic alliance.

Laboratory testing is likely to identify the cause of chronic fatigue in only about 5% of cases. Beyond a few standard screening tests, laboratory evaluation should be guided by the history and physical examination; extensive testing is likely to lead to incidental findings that require explanation and unnecessary follow-up investigation, and should be avoided in lieu of frequent clinical follow-up. A reasonable approach to screening includes a complete blood count with differential (to screen for anemia, infection, and malignancy), electrolytes (including sodium, potassium, and calcium), glucose, renal function, liver function, and thyroid function. Testing for HIV and adrenal function can also be considered. Published guidelines for chronic fatigue syndrome also recommend an erythrocyte sedimentation rate (ESR) as part of the evaluation for mimics, but unless the value is very high, such nonspecific testing in the absence of other features is unlikely to clarify the situation. Routine screening with an antinuclear antibody (ANA) test is also unlikely to be informative in isolation and is frequently positive at low titers in otherwise healthy adults. Additional unfocused studies, such as whole-body imaging scans, are usually not indicated; in addition to their inconvenience, potential risk, and cost, they often reveal unrelated incidental findings that can prolong the workup unnecessarily.

TREATMENT

Fatigue

The first priority is to address the underlying disorder or disorders that account for fatigue, because this can be curative in select contexts and palliative in others. Unfortunately, in many chronic

illnesses, fatigue may be refractory to traditional disease-modifying therapies, but it is nevertheless important in such cases to evaluate for other potential contributors because the cause may be multifactorial. Antidepressants (**Chap. 452**) may be helpful for treatment of chronic fatigue when symptoms of depression are present and are generally most effective as part of a multimodal approach. However, antidepressants can also cause fatigue and should be discontinued if they are not clearly effective. Cognitive-behavioral therapy has also been demonstrated to be helpful in ME/CFS as well as cancer-associated fatigue. Both cognitive-behavioral therapy and graded exercise therapy, in which physical exercise, most typically walking, is gradually increased with attention to target heart rates to avoid overexertion, were shown to modestly improve walking times and self-reported fatigue measures when compared to standard medical care in patients in the United Kingdom with chronic fatigue. These benefits were maintained after a median follow-up of 2.5 years. Psychostimulants such as amphetamines, modafinil, and armodafinil can help increase alertness and concentration and reduce excessive daytime sleepiness in certain clinical contexts, which may in turn help with symptoms of fatigue in a minority of patients, but they have generally proven to be unhelpful in randomized trials for treating fatigue in posttraumatic brain injury, Parkinson's disease, cancer, and MS. In patients with low vitamin D status, vitamin D replacement may lead to improvement in fatigue.

Development of more effective therapy for fatigue is hampered by limited knowledge of the biologic basis of this symptom, including how fatigue is detected and registered in the nervous system. Proinflammatory cytokines, such as interleukin 1 α and 1 β and tumor necrosis factor α , might mediate fatigue in some patients. While preliminary studies of biologic therapies that inhibit cytokines have suggested a benefit against fatigue in some patients with inflammatory conditions, this approach has largely not led to improvement in clinical trials that focused on fatigue as the primary endpoint. Nonetheless, specific targeting with cytokine antagonists could represent a possible future approach for some patients.

■ PROGNOSIS

Acute fatigue significant enough to require medical evaluation is more likely to lead to an identifiable medical, neurologic, or psychiatric cause than is unexplained chronic fatigue. Evaluation of unexplained chronic fatigue most commonly leads to diagnosis of a psychiatric condition or remains unexplained. Identification of a previously undiagnosed serious or life-threatening culprit etiology is rare, even with longitudinal follow-up of patients with unexplained chronic fatigue. Complete resolution is uncommon, at least over the short term, but multidisciplinary treatment approaches can lead to symptomatic improvements that substantially improve quality of life.

■ FURTHER READING

- BASU N et al: Fatigue is associated with excess mortality in the general population: Results from the EPIC-Norfolk study. *BMC Med* 14:122, 2016.
- DUKES JC et al: Approach to fatigue: Best practice. *Med Clin North Am* 105:137, 2021.
- ROERINK ME et al: Interleukin-1 as a mediator of fatigue in disease: A narrative review. *J Neuroinflammation* 14:16, 2017.
- SHARPE M et al: Rehabilitative treatments for chronic fatigue syndrome: Long-term follow-up from the PACE trial. *Lancet Psychiatry* 2:1067, 2015.
- WHITE PD et al: Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): A randomised trial. *Lancet* 377:823, 2011.

24

Neurologic Causes of Weakness and Paralysis

Stephen L. Hauser



Normal motor function involves integrated muscle activity that is modulated by the activity of the cerebral cortex, basal ganglia, cerebellum, red nucleus, brainstem reticular formation, lateral vestibular nucleus, and spinal cord. Motor system dysfunction leads to weakness or paralysis, discussed in this chapter, or to ataxia (Chap. 439) or abnormal movements (Chap. 436). *Weakness* is a reduction in the power that can be exerted by one or more muscles. It must be distinguished from increased *fatigability* (i.e., the inability to sustain the performance of an activity that should be normal for a person of the same age, sex, and size), limitation in function due to pain or articular stiffness, or impaired motor activity because severe *proprioceptive sensory loss* prevents adequate feedback information about the direction and power of movements. It is also distinct from *bradykinesia* (in which increased time is required for full power to be exerted) and *apraxia*, a disorder of planning and initiating a skilled or learned movement unrelated to a significant motor or sensory deficit (Chap. 30).

Paralysis or the suffix “-plegia” indicates weakness so severe that a muscle cannot be contracted at all, whereas *paresis* refers to less severe weakness. The prefix “hemi-” refers to one-half of the body, “para-” to both legs, and “quadri-” to all four limbs.

The *distribution* of weakness helps to localize the underlying lesion. Weakness from involvement of upper motor neurons occurs particularly in the extensors and abductors of the upper limb and the flexors of the lower limb. Lower motor neuron weakness depends on whether involvement is at the level of the anterior horn cells, nerve root, limb plexus, or peripheral nerve—only muscles supplied by the affected structure are weak. Myopathic weakness is generally most marked in proximal muscles. Weakness from impaired neuromuscular transmission has no specific pattern of involvement.

Weakness often is accompanied by other neurologic abnormalities that help indicate the site of the responsible lesion (Table 24-1).

Tone is the resistance of a muscle to passive stretch. Increased tone may be of several types. *Spasticity* is the increase in tone associated with disease of upper motor neurons. It is velocity dependent, has a sudden release after reaching a maximum (the “clasp-knife” phenomenon), and predominantly affects the antigravity muscles (i.e., upper-limb flexors and lower-limb extensors). *Rigidity* is hypertonia that is present throughout the range of motion (a “lead pipe” or “plastic” stiffness) and affects flexors and extensors equally; it sometimes has a cogwheel quality that is enhanced by voluntary movement of the contralateral limb (reinforcement). Rigidity occurs with certain extrapyramidal disorders, such as Parkinson’s disease. *Paratonia* (or *gegenhalten*) is increased tone that varies irregularly in a manner seemingly related to the degree of relaxation, is present throughout the range of motion, and affects flexors and extensors equally; it usually results from disease of the frontal lobes. Weakness with *decreased tone (flaccidity)* or normal tone occurs with disorders of *motor units*. A motor unit consists of a single lower motor neuron and all the muscle fibers that it innervates.

Muscle bulk generally is not affected by upper motor neuron lesions, although mild disuse atrophy eventually may occur. By contrast, atrophy is often conspicuous when a lower motor neuron lesion is responsible for weakness and also may occur with advanced muscle disease.

Muscle stretch (tendon) reflexes are usually increased with upper motor neuron lesions but may be decreased or absent for a variable period immediately after onset of an acute lesion. Hyperreflexia is usually—but not invariably—accompanied by loss of *cutaneous reflexes* (such as superficial abdominals; Chap. 422) and, in particular, by an extensor plantar (Babinski) response. The muscle stretch reflexes are depressed with lower motor neuron lesions directly involving specific reflex arcs. They generally are preserved in patients with myopathic weakness except in advanced stages, when they sometimes are attenuated. In disorders of the neuromuscular junction, reflex responses may be affected by preceding voluntary activity of affected muscles; such activity may lead to enhancement of initially depressed reflexes in Lambert-Eaton myasthenic syndrome and, conversely, to depression of initially normal reflexes in myasthenia gravis (Chap. 448).

The distinction of *neuropathic* (lower motor neuron) from *myopathic* weakness is sometimes difficult clinically, although distal weakness is likely to be neuropathic, and symmetric proximal weakness myopathic. *Fasciculations* (visible or palpable twitches within a muscle due to the spontaneous discharge of a motor unit) and early atrophy indicate that weakness is myopathic.

PATHOGENESIS

Upper Motor Neuron Weakness Lesions of the upper motor neurons or their descending axons to the spinal cord (Fig. 24-1) produce weakness through decreased activation of lower motor neurons. In general, distal muscle groups are affected more severely than proximal ones, and axial movements are spared unless the lesion is severe and bilateral. Spasticity is typical but may not be present acutely. Rapid repetitive movements are slowed and coarse, but normal rhythmicity is maintained. With corticobulbar involvement, weakness occurs in the lower face and tongue; extraocular, upper facial, pharyngeal, and jaw muscles are typically spared. Bilateral corticobulbar lesions produce a *pseudobulbar palsy*: dysarthria, dysphagia, dysphonia, and emotional lability accompany bilateral facial weakness and a brisk jaw jerk. Electromyogram (EMG) (Chap. 446) shows that with weakness of the upper motor neuron type, motor units have a diminished maximal discharge frequency.

Lower Motor Neuron Weakness This pattern results from disorders of lower motor neurons in the brainstem motor nuclei and the anterior horn of the spinal cord or from dysfunction of the axons of these neurons as they pass to skeletal muscle (Fig. 24-2). Weakness is due to a decrease in the number of muscle fibers that can be activated through a loss of a motor neurons or disruption of their connections to muscle. Loss of γ motor neurons does not cause weakness but decreases tension on the muscle spindles, which decreases muscle tone and attenuates the stretch reflexes. An absent stretch reflex suggests involvement of spindle afferent fibers.

When a motor unit becomes diseased, especially in anterior horn cell diseases, it may discharge spontaneously, producing *fasciculations*. When α motor neurons or their axons degenerate, the denervated muscle fibers also may discharge spontaneously. These single muscle

TABLE 24-1 Signs That Distinguish the Origin of Weakness

SIGN	UPPER MOTOR NEURON	LOWER MOTOR NEURON	MYOPATHIC	PSYCHOGENIC
Atrophy	None	Severe	Mild	None
Fasciculations	None	Common	None	None
Tone	Spastic	Decreased	Normal/decreased	Variable/paratonia
Distribution of weakness	Pyramidal/regional	Distal/segmental	Proximal	Variable/inconsistent with daily activities
Muscle stretch reflexes	Hyperactive	Hypoactive/absent	Normal/hypoactive	Normal
Babinski sign	Present	Absent	Absent	Absent

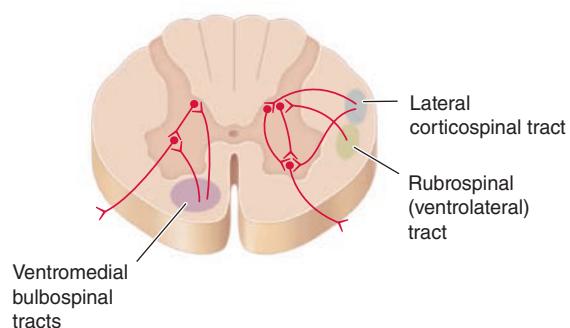
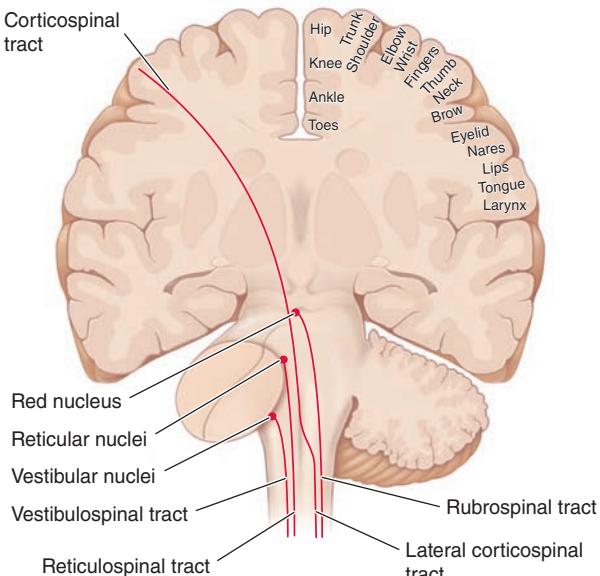


FIGURE 24-1 The corticospinal and bulbospinal upper motor neuron pathways. Upper motor neurons have their cell bodies in layer V of the primary motor cortex (the precentral gyrus, or Brodmann area 4) and in the premotor and supplemental motor cortex (area 6). The upper motor neurons in the primary motor cortex are somatotopically organized (*right side of figure*). Axons of the upper motor neurons descend through the subcortical white matter and the posterior limb of the internal capsule. Axons of the pyramidal or corticospinal system descend through the brainstem in the cerebral peduncle of the midbrain, the basis pontis, and the medullary pyramids. At the cervicomедullary junction, most corticospinal axons decussate into the contralateral corticospinal tract of the lateral spinal cord, but 10–30% remain ipsilateral in the anterior spinal cord. Corticospinal neurons synapse on premotor interneurons, but some—especially in the cervical enlargement and those connecting with motor neurons to distal limb muscles—make direct monosynaptic connections with lower motor neurons. They innervate most densely the lower motor neurons of hand muscles and are involved in the execution of learned, fine movements. Corticobulbar neurons are similar to corticospinal neurons but innervate brainstem motor nuclei. Bulbospinal upper motor neurons influence strength and tone but are not part of the pyramidal system. The descending ventromedial bulbospinal pathways originate in the tectum of the midbrain (tectospinal pathway), the vestibular nuclei (vestibulospinal pathway), and the reticular formation (reticulospinal pathway). These pathways influence axial and proximal muscles and are involved in the maintenance of posture and integrated movements of the limbs and trunk. The descending ventrolateral bulbospinal pathways, which originate predominantly in the red nucleus (rubrospinal pathway), facilitate distal limb muscles. The bulbospinal system sometimes is referred to as the extrapyramidal upper motor neuron system. In all figures, nerve cell bodies and axon terminals are shown, respectively, as closed circles and forks.

fiber discharges, or *fibrillation potentials*, cannot be seen but can be recorded with EMG. Weakness leads to delayed or reduced recruitment of motor units, with fewer than normal activated at a particular discharge frequency.

Neuromuscular Junction Weakness Disorders of the neuromuscular junction produce weakness of variable degree and distribution. The number of muscle fibers that are activated varies over time, depending on the state of rest of the neuromuscular junctions. Strength

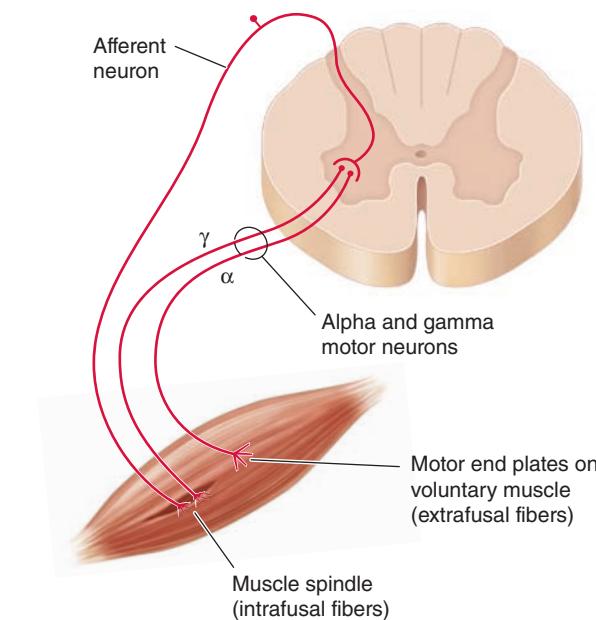


FIGURE 24-2 Lower motor neurons are divided into α and γ types. The larger α motor neurons are more numerous and innervate the extrafusal muscle fibers of the motor unit. Loss of α motor neurons or disruption of their axons produces lower motor neuron weakness. The smaller, less numerous γ motor neurons innervate the intrafusal muscle fibers of the muscle spindle and contribute to normal tone and stretch reflexes. The α motor neuron receives direct excitatory input from corticomotoneurons and primary muscle spindle afferents. The α and γ motor neurons also receive excitatory input from other descending upper motor neuron pathways, segmental sensory inputs, and interneurons. The α motor neurons receive direct inhibition from Renshaw cell interneurons, and other interneurons indirectly inhibit the α and γ motor neurons. A muscle stretch (tendon) reflex requires the function of all the illustrated structures. A tap on a tendon stretches muscle spindles (which are tonically activated by γ motor neurons) and activates the primary spindle afferent neurons. These neurons stimulate the α motor neurons in the spinal cord, producing a brief muscle contraction, which is the familiar tendon reflex.

is influenced by preceding activity of the affected muscle. In myasthenia gravis, for example, sustained or repeated contractions of affected muscle decline in strength despite continuing effort (Chap. 440). Thus, fatigable weakness is suggestive of disorders of the neuromuscular junction, which cause functional loss of muscle fibers due to failure of their activation.

Myopathic Weakness Myopathic weakness is produced by a decrease in the number or contractile force of muscle fibers activated within motor units. With muscular dystrophies, inflammatory myopathies, or myopathies with muscle fiber necrosis, the number of muscle fibers is reduced within many motor units. On EMG, the size of each motor unit action potential is decreased, and motor units must be recruited more rapidly than normal to produce the desired power. Some myopathies produce weakness through loss of contractile force of muscle fibers or through relatively selective involvement of type II (fast) fibers. These myopathies may not affect the size of individual motor unit action potentials and are detected by a discrepancy between the electrical activity and force of a muscle.

Psychogenic Weakness Weakness may occur without a recognizable organic basis. It tends to be variable, inconsistent, and with a pattern of distribution that cannot be explained on a neuroanatomic basis. On formal testing, antagonists may contract when the patient is supposedly activating the agonist muscle. The severity of weakness is out of keeping with the patient's daily activities.

DISTRIBUTION OF WEAKNESS

Hemiparesis Hemiparesis results from an upper motor neuron lesion above the midcervical spinal cord; most such lesions are above the foramen magnum. The presence of other neurologic deficits helps localize the lesion. Thus language disorders, for example, point to a

cortical lesion. Homonymous visual field defects reflect either a cortical or a subcortical hemispheric lesion. A “pure motor” hemiparesis of the face, arm, and leg often is due to a small, discrete lesion in the posterior limb of the internal capsule, cerebral peduncle in the midbrain, or upper pons. Some brainstem lesions produce “crossed paralyses,” consisting of ipsilateral cranial nerve signs and contralateral hemiparesis (Chap. 426). The absence of cranial nerve signs or facial weakness suggests that a hemiparesis is due to a lesion in the high cervical spinal cord, especially if associated with Brown-Séquard syndrome, consisting of loss of joint position and vibration sense on the side of the weakness, and loss of pain and temperature sense on the opposite side (Chap. 442).

Acute or episodic hemiparesis usually results from focal structural lesions, particularly vascular etiologies, rapidly expanding lesions, or an inflammatory process. **Subacute hemiparesis** that evolves over days or weeks may relate to subdural hematoma, infectious or inflammatory disorders (e.g., cerebral abscess, fungal granuloma or meningitis, parasitic infection, multiple sclerosis, sarcoidosis), or primary or metastatic neoplasms. AIDS may present with subacute hemiparesis due to toxoplasmosis or primary central nervous system (CNS) lymphoma. **Chronic hemiparesis** that evolves over months usually is due to a neoplasm or vascular malformation, a chronic subdural hematoma, or a degenerative disease.

Investigation of hemiparesis (Fig. 24-3) of acute origin usually starts with a CT scan of the brain and laboratory studies. If the CT is normal, or in subacute or chronic cases of hemiparesis, MRI of the brain and/or cervical spine (including the foramen magnum) is performed, depending on the clinical accompaniments.

Paraparesis **Acute paraparesis** is caused most commonly by an intraspinal lesion, but its spinal origin may not be recognized initially if the legs are flaccid and areflexic. Usually, however, there is sensory loss in the legs with an upper level on the trunk; a dissociated sensory loss (loss of pain and temperature but not touch, position, and vibration sense) suggestive of a central cord syndrome; or hyperreflexia in the legs with normal reflexes in the arms (Chap. 442). Imaging the

spinal cord (Fig. 24-3) may reveal compressive lesions, infarction (proprioception usually is spared), arteriovenous fistulas or other vascular anomalies, or transverse myelitis (Chap. 442).

Diseases of the cerebral hemispheres that produce acute paraparesis include anterior cerebral artery ischemia (shoulder shrug also is affected), superior sagittal sinus or cortical venous thrombosis, and acute hydrocephalus.

Paraparesis may also result from a cauda equina syndrome, for example, after trauma to the low back, a midline disk herniation, or an intraspinal tumor. The sphincters are commonly affected, whereas hip flexion often is spared, as is sensation over the anterolateral thighs. Rarely, paraparesis is caused by a rapidly evolving anterior horn cell disease (such as poliovirus or West Nile virus infection), peripheral neuropathy (such as Guillain-Barré syndrome; Chap. 447), or myopathy (Chap. 449).

Subacute or chronic spastic paraparesis is caused by upper motor neuron disease. When associated with lower-limb sensory loss and sphincter involvement, a chronic spinal cord disorder should be considered (Chap. 442). If hemispheric signs are present, a parasagittal meningioma or chronic hydrocephalus is likely. The absence of spasticity in a long-standing paraparesis suggests a lower motor neuron or myopathic etiology.

Investigations typically begin with spinal MRI, but when upper motor neuron signs are associated with drowsiness, confusion, seizures, or other hemispheric signs, brain MRI should also be performed, sometimes as the initial investigation. Electrophysiologic studies are diagnostically helpful when clinical findings suggest an underlying neuromuscular disorder.

Quadriplegia or Generalized Weakness Generalized weakness may be due to disorders of the CNS or the motor unit. Although the terms often are used interchangeably, **quadripareisis** is commonly used when an upper motor neuron cause is suspected, and **generalized weakness** is used when a disease of the motor units is likely. Weakness from CNS disorders usually is associated with changes in consciousness or cognition and accompanied by spasticity, hyperreflexia, and sensory disturbances.

Most neuromuscular causes of generalized weakness are associated with normal mental function, hypotonia, and hypoactive muscle stretch reflexes. The major causes of intermittent weakness are listed in Table 24-2. A patient with generalized fatigability without objective weakness may have chronic fatigue syndrome (Chap. 450).

ACUTE QUADRIPLAESIS Quadripareisis with onset over minutes may result from disorders of upper motor neurons (such as from anoxia, hypotension, brainstem or cervical cord ischemia, trauma, and systemic metabolic abnormalities) or muscle (electrolyte disturbances, certain inborn errors of muscle energy metabolism, toxins, and periodic paralyses). Onset over hours to weeks may, in addition to these disorders, be due to lower motor neuron disorders such as Guillain-Barré syndrome (Chap. 447).

In obtunded patients, evaluation begins with a CT or MRI scan of the brain. If upper motor neuron signs are present but the patient is alert, the initial test is usually an MRI of the cervical cord. If weakness is lower motor neuron, myopathic, or uncertain in origin, the clinical approach begins with blood studies to determine the level of muscle enzymes and electrolytes and with EMG and nerve conduction studies.

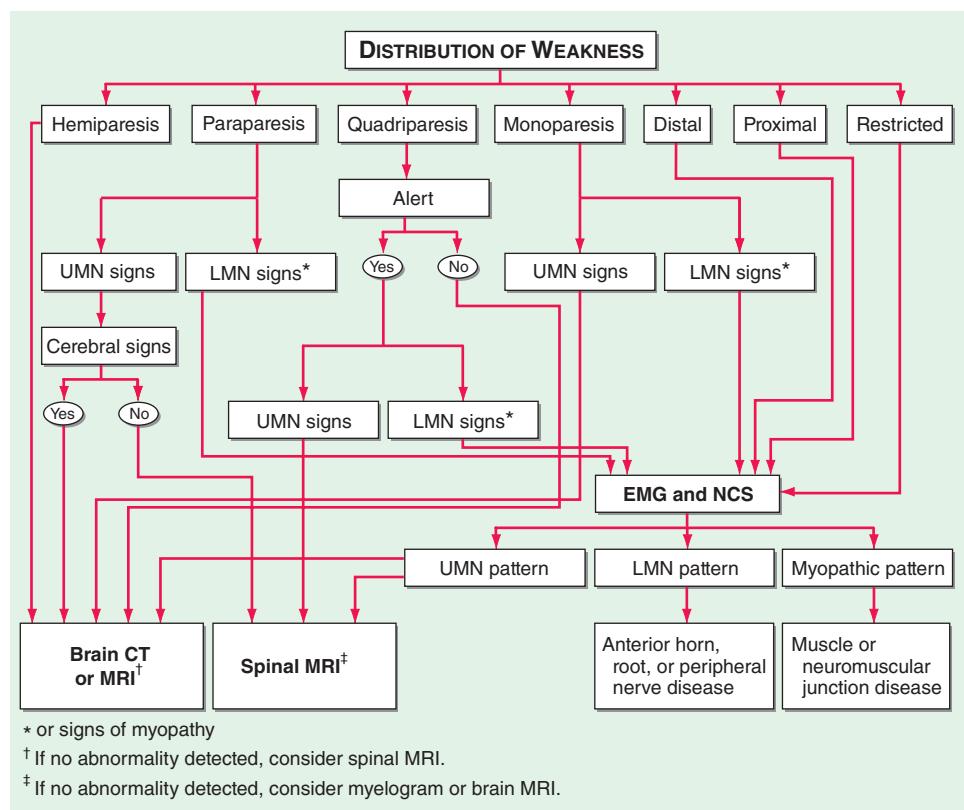


FIGURE 24-3 An algorithm for the initial workup of a patient with weakness. CT, computed tomography; EMG, electromyography; LMN, lower motor neuron; MRI, magnetic resonance imaging; NCS, nerve conduction studies; UMN, upper motor neuron.

TABLE 24-2 Causes of Episodic Generalized Weakness

1. Electrolyte disturbances, e.g., hypokalemia, hyperkalemia, hypercalcemia, hypernatremia, hyponatremia, hypophosphatemia, hypermagnesemia
2. Muscle disorders
 - a. Channelopathies (periodic paralyses)
 - b. Metabolic defects of muscle (impaired carbohydrate or fatty acid utilization; abnormal mitochondrial function)
3. Neuromuscular junction disorders
 - a. Myasthenia gravis
 - b. Lambert-Eaton myasthenic syndrome
4. Central nervous system disorders
 - a. Transient ischemic attacks of the brainstem
 - b. Transient global cerebral ischemia
 - c. Multiple sclerosis
5. Lack of voluntary effort
 - a. Anxiety
 - b. Pain or discomfort
 - c. Somatization disorder

SUBACUTE OR CHRONIC QUADRIPIARESIS Quadripareisis due to upper motor neuron disease may develop over weeks to years from chronic myelopathies, multiple sclerosis, brain or spinal tumors, chronic subdural hematomas, and various metabolic, toxic, and infectious disorders. It may also result from lower motor neuron disease, a chronic neuropathy (in which weakness is often most profound distally), or myopathic weakness (typically proximal).

When quadripareisis develops acutely in obtunded patients, evaluation begins with a CT scan of the brain. If upper motor neuron signs have developed acutely but the patient is alert, the initial test is usually an MRI of the cervical cord. When onset has been gradual, disorders of the cerebral hemispheres, brainstem, and cervical spinal cord can usually be distinguished clinically, and imaging is directed first at the clinically suspected site of pathology. If weakness is lower motor neuron, myopathic, or uncertain in origin, laboratory studies can determine the levels of muscle enzymes and electrolytes, and EMG and nerve conduction studies help to localize the pathologic process (**Chap. 449**).

Monoparesis Monoparesis usually is due to lower motor neuron disease, with or without associated sensory involvement. Upper motor neuron weakness occasionally presents as a monoparesis of distal and nonantigravity muscles. Myopathic weakness rarely is limited to one limb.

ACUTE MONOPARESIS If weakness is predominantly distal and of upper motor neuron type and is not associated with sensory impairment or pain, focal cortical ischemia is likely (**Chap. 427**); diagnostic possibilities are similar to those for acute hemiparesis. Sensory loss and pain usually accompany acute lower motor neuron weakness; the weakness commonly localizes to a single nerve root or peripheral nerve, but occasionally reflects plexus involvement. If lower motor neuron weakness is likely, evaluation begins with EMG and nerve conduction studies.

SUBACUTE OR CHRONIC MONOPARESIS Weakness and atrophy that develop over weeks or months are usually of lower motor neuron origin. When associated with sensory symptoms, a peripheral cause (nerve, root, or plexus) is likely; otherwise, anterior horn cell disease should be considered. In either case, an electrodiagnostic study is indicated. If weakness is of the upper motor neuron type, a discrete cortical (precentral gyrus) or cord lesion may be responsible, and appropriate imaging is performed.

Distal Weakness Involvement of two or more limbs distally suggests lower motor neuron or peripheral nerve disease. Acute distal lower-limb weakness results occasionally from an acute toxic polyneuropathy or cauda equina syndrome. Distal symmetric weakness usually develops over weeks, months, or years and, when associated with numbness, is due to peripheral neuropathy (**Chap. 446**). Anterior horn

cell disease may begin distally but is typically asymmetric and without accompanying numbness (**Chap. 437**). Rarely, myopathies present with distal weakness (**Chap. 449**). Electrodiagnostic studies help localize the disorder (Fig. 24-3).

Proximal Weakness Myopathy often produces symmetric weakness of the pelvic or shoulder girdle muscles (**Chap. 449**). Diseases of the neuromuscular junction, such as myasthenia gravis (**Chap. 448**), may present with symmetric proximal weakness often associated with ptosis, diplopia, or bulbar weakness and fluctuate in severity during the day. In anterior horn cell disease, proximal weakness is usually asymmetric, but it may be symmetric especially in genetic forms. Numbness does not occur with any of these diseases. The evaluation usually begins with determination of the serum creatine kinase level and electrophysiologic studies.

Weakness in a Restricted Distribution Weakness may not fit any of these patterns, being limited, for example, to the extraocular, hemifacial, bulbar, or respiratory muscles. If it is unilateral, restricted weakness usually is due to lower motor neuron or peripheral nerve disease, such as in a facial palsy. Weakness of part of a limb is commonly due to a peripheral nerve lesion such as an entrapment neuropathy. Relatively symmetric weakness of extraocular or bulbar muscles frequently is due to a myopathy (**Chap. 449**) or neuromuscular junction disorder (**Chap. 448**). Bilateral facial palsy with areflexia suggests Guillain-Barré syndrome (**Chap. 447**). Worsening of relatively symmetric weakness with fatigue is characteristic of neuromuscular junction disorders. Asymmetric bulbar weakness usually is due to motor neuron disease. Weakness limited to respiratory muscles is uncommon and usually is due to motor neuron disease, myasthenia gravis, or polymyositis/dermatomyositis (**Chap. 365**).

ACKNOWLEDGMENT

The editors acknowledge the contributions of Michael J. Aminoff to earlier editions of this chapter.

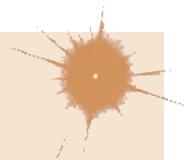
■ FURTHER READING

- BRAZIS P et al: *Localization in Clinical Neurology*, 7th ed. Philadelphia, Lippincott William & Wilkins, 2016.
 CAMPBELL WW, BAROHN RJ: *DeJong's The Neurological Examination*, 8th ed. Philadelphia, Lippincott William & Wilkins, 2019.
 GUARANTORS OF BRAIN: *Aids to the Examination of the Peripheral Nervous System*, 4th ed. Edinburgh, Saunders, 2000.

25

Numbness, Tingling, and Sensory Loss

Stephen L. Hauser



Normal somatic sensation reflects a continuous monitoring process, little of which reaches consciousness under ordinary conditions. By contrast, disordered sensation, particularly when experienced as painful, is alarming and dominates the patient's attention. Physicians should be able to recognize abnormal sensations by how they are described, know their type and likely site of origin, and understand their implications. **Pain is considered separately in Chap. 13.**

■ POSITIVE AND NEGATIVE SYMPTOMS

Abnormal sensory symptoms can be divided into two categories: positive and negative. The prototypical positive symptom is tingling (pins and needles); other positive sensory phenomena include itch and altered sensations that are described as pricking, bandlike, lightning-like shooting feelings (lancinations), aching, knifelike, twisting, drawing,

pulling, tightening, burning, searing, electrical, or raw feelings. Such symptoms are often painful.

Positive phenomena usually result from trains of impulses generated at sites of lowered threshold or heightened excitability along a peripheral or central sensory pathway. The nature and severity of the abnormal sensation depend on the number, rate, timing, and distribution of ectopic impulses and the type and function of nervous tissue in which they arise. Because positive phenomena represent excessive activity in sensory pathways, they are not necessarily associated with a sensory deficit (loss) on examination.

Negative phenomena represent loss of sensory function and are characterized by diminished or absent feeling that often is experienced as numbness and by abnormal findings on sensory examination. In disorders affecting peripheral sensation, at least one-half of the afferent axons innervating a particular site are probably lost or functionless before a sensory deficit can be demonstrated by clinical examination. If the rate of loss is slow, however, lack of cutaneous feeling may be unnoticed by the patient and difficult to demonstrate on examination, even though few sensory fibers are functioning; if it is rapid, both positive and negative phenomena are usually conspicuous. Subclinical degrees of sensory dysfunction may be revealed by sensory nerve conduction studies or somatosensory-evoked potentials.

Whereas sensory symptoms may be either positive or negative, sensory signs on examination are always a measure of negative phenomena.

TERMINOLOGY

Paresthesias and dysesthesias are general terms used to denote positive sensory symptoms. The term *paresthesias* typically refers to tingling or pins-and-needles sensations but may include a wide variety of other abnormal sensations, except pain; it sometimes implies that the abnormal sensations are perceived spontaneously. The more general term *dysesthesias* denotes all types of abnormal sensations, including painful ones, regardless of whether a stimulus is evident.

Another set of terms refers to sensory abnormalities found on examination. *Hypesthesia* or *hypoesthesia* refers to a reduction of cutaneous sensation to a specific type of testing such as pressure, light touch, and warm or cold stimuli; *anesthesia*, to a complete absence of skin sensation to the same stimuli plus pinprick; and *hypalgesia* or *analgesia*, to reduced or absent pain perception (nociception). *Hyperesthesia* means pain or increased sensitivity in response to touch. Similarly, *allodynia* describes the situation in which a nonpainful stimulus, once perceived, is experienced as painful, even excruciating. An example is elicitation of a painful sensation by application of a vibrating tuning fork. *Hyperalgesia* denotes severe pain in response to a mildly noxious stimulus, and *hyperpathia*, a broad term, encompasses all the phenomena described by hyperesthesia, allodynia, and hyperalgesia. With hyperpathia, the threshold for a sensory stimulus is increased and perception is delayed, but once felt, it is unduly painful.

Disorders of deep sensation arising from muscle spindles, tendons, and joints affect proprioception (position sense). Manifestations include imbalance (particularly with eyes closed or in the dark), clumsiness of precision movements, and unsteadiness of gait, which are referred to collectively as *sensory ataxia*. Other findings on examination usually, but not invariably, include reduced or absent joint position and vibratory sensibility and absent deep tendon reflexes in the affected limbs. The Romberg sign is positive, which means that the patient sways markedly or topples when asked to stand with feet close together and eyes closed. In severe states of deafferentation involving deep sensation, the patient cannot walk or stand unaided or even sit unsupported. Continuous involuntary movements (*pseudoathetosis*) of the outstretched hands and fingers occur, particularly with eyes closed.

ANATOMY OF SENSATION

Cutaneous receptors are classified by the type of stimulus that optimally excites them. They consist of naked nerve endings (nociceptors, which respond to tissue-damaging stimuli, and thermoreceptors, which respond to noninjurious thermal stimuli) and encapsulated terminals (several types of mechanoreceptor, activated by physical

deformation of the skin or stretch of muscles). Each type of receptor has its own set of sensitivities to specific stimuli, size and distinctness of receptive fields, and adaptational qualities.

Afferent peripheral nerve fibers conveying somatosensory information from the limbs and trunk traverse the dorsal roots and enter the dorsal horn of the spinal cord (Fig. 25-1), the cell bodies of first-order neurons are located in the dorsal root ganglia (DRG). In an analogous fashion, sensations from the face and head are conveyed through the trigeminal system (Fig. 441-2). Once fiber tracts enter the spinal cord, the polysynaptic projections of the smaller fibers (unmyelinated and small myelinated), which subserve mainly nociception, itch, temperature sensibility, and touch, cross and ascend in the opposite anterior and lateral columns of the spinal cord, through the brainstem, to the ventral posterolateral (VPL) nucleus of the thalamus and ultimately project to the postcentral gyrus of the parietal cortex and other cortical areas (Chap. 13). This is the *spinothalamic pathway* or *anterolateral system*. The larger fibers, which subserve tactile and position sense and kinesthesia, project rostrally in the posterior and posterolateral columns on the same side of the spinal cord and make their first synapse in the gracile or cuneate nucleus of the lower medulla. Axons of second-order neurons decussate and ascend in the medial lemniscus located medially in the medulla and in the tegmentum of the pons and midbrain and synapse in the VPL nucleus; third-order neurons project to parietal cortex as well as to other cortical areas. This large-fiber system is referred to as the *posterior column-medial lemniscal pathway* (lemniscal, for short). Although the fiber types and functions that make up the spinothalamic and lemniscal systems are relatively well known, many other fibers, particularly those associated with touch, pressure, and position sense, ascend in a diffusely distributed pattern both ipsilaterally and contralaterally in the anterolateral quadrants of the spinal cord. This explains why a complete lesion of the posterior columns of the spinal cord may be associated with little sensory deficit on examination.

APPROACH TO THE PATIENT

Clinical Examination of Sensation

The main components of the sensory examination are tests of primary sensation (pain, touch, vibration, joint position, and thermal sensation) (Table 25-1). The examiner must depend on patient responses, and this complicates interpretation. Further, examination may be limited in some patients. In a stuporous patient, for example, sensory examination is reduced to observing the briskness of withdrawal in response to a pinch or another noxious stimulus. Comparison of responses on the two sides of the body is essential. In an alert but uncooperative patient, it may not be possible to examine cutaneous sensation, but some idea of proprioceptive function may be gained by noting the patient's best performance of movements requiring balance and precision.

In patients with sensory complaints, testing should begin in the center of the affected region and proceed radially until sensation is perceived as normal. The distribution of any abnormality is defined and compared to root and peripheral nerve territories (Figs. 25-2 and 25-3). Some patients present with sensory symptoms that do not fit an anatomic localization and are accompanied by either no abnormalities or gross inconsistencies on examination. The examiner should consider in such cases the possibility of a psychologic cause (see "Psychogenic Symptoms," below). Sensory examination of a patient who has no neurologic complaints can be brief and consist of pinprick, touch, and vibration testing in the hands and feet plus evaluation of stance and gait, including the Romberg maneuver (Chap. V6). Evaluation of stance and gait also tests the integrity of motor and cerebellar systems.

PRIMARY SENSATION

The sense of pain usually is tested with a clean pin, which is then discarded. The patient is asked to close the eyes and focus on the pricking or unpleasant quality of the stimulus, not just the pressure

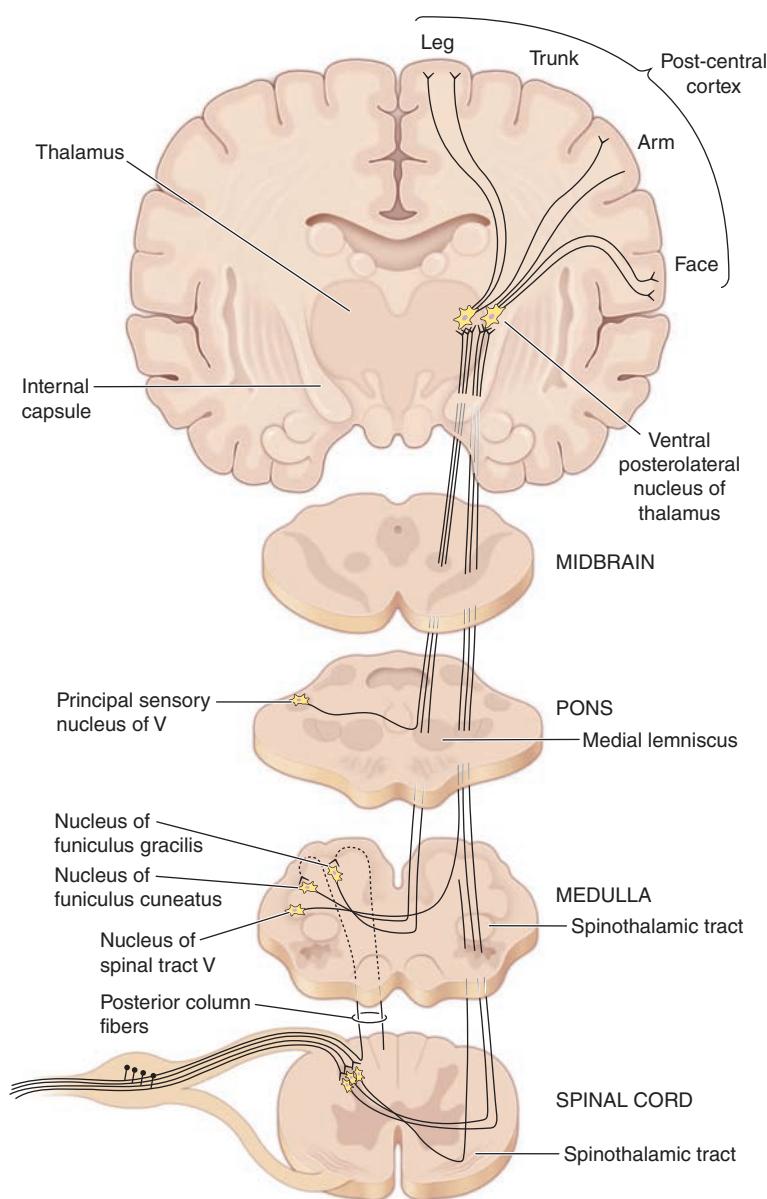


FIGURE 25-1 The main somatosensory pathways. The spinothalamic tract (pain, thermal sense) and the posterior column–lemniscal system (touch, pressure, joint position) are shown. Offshoots from the ascending anterolateral fasciculus (spinothalamic tract) to nuclei in the medulla, pons, and mesencephalon and nuclear terminations of the tract are indicated. (Reproduced with permission from AH Ropper, MA Samuels: Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009.)

or touch sensation elicited. Areas of hypalgesia should be mapped by proceeding radially from the most hypalgesic site. Temperature sensation to both hot and cold is best tested with small containers filled with water of the desired temperature. An alternative way to test cold sensation is to touch a metal object, such as a tuning fork

at room temperature, to the skin. For testing warm temperatures, the tuning fork or another metal object may be held under warm water of the desired temperature and then used. The appreciation of both cold and warmth should be tested because different receptors respond to each. Touch usually is tested with a wisp of cotton,

TABLE 25-1 Testing Primary Sensation

SENSE	TEST DEVICE	ENDINGS ACTIVATED	FIBER SIZE MEDIATING	CENTRAL PATHWAY
Pain	Pinprick	Cutaneous nociceptors	Small	SpTh, also D
Temperature, heat	Warm metal object	Cutaneous thermoreceptors for hot	Small	SpTh
Temperature, cold	Cold metal object	Cutaneous thermoreceptors for cold	Small	SpTh
Touch	Cotton wisp, fine brush	Cutaneous mechanoreceptors, also naked endings	Large and small	Lem, also D and SpTh
Vibration	Tuning fork, 128 Hz	Mechanoreceptors, especially pacinian corpuscles	Large	Lem, also D
Joint position	Passive movement of specific joints	Joint capsule and tendon endings, muscle spindles	Large	Lem, also D

Abbreviations: D, diffuse ascending projections in ipsilateral and contralateral anterolateral columns; Lem, posterior column and lemniscal projection, ipsilateral; SpTh, spinothalamic projection, contralateral.

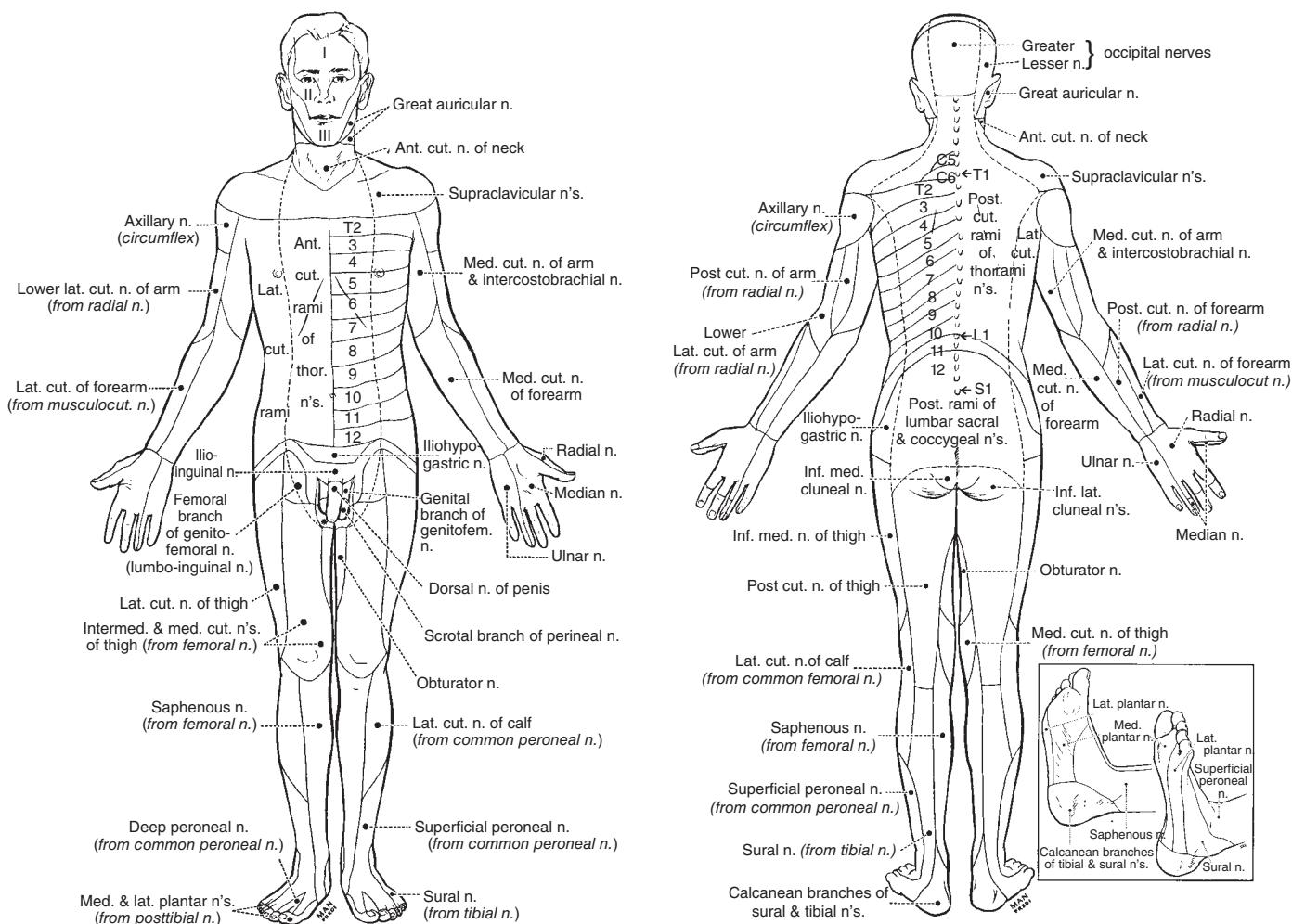


FIGURE 25-2 The cutaneous fields of peripheral nerves. (Reproduced with permission from W Haymaker, B Woodhall: *Peripheral Nerve Injuries*, 2nd ed. Philadelphia, Saunders, 1953.)

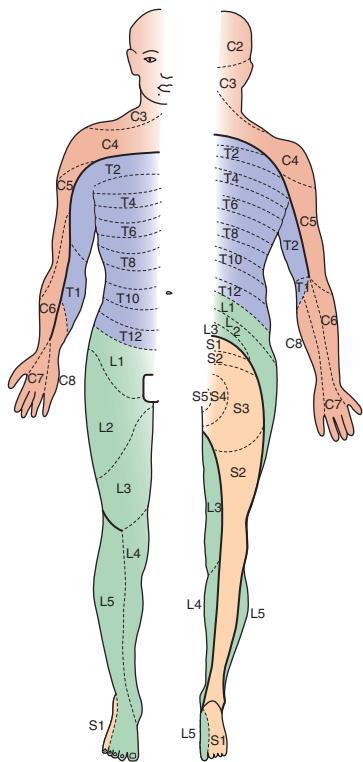


FIGURE 25-3 Distribution of the sensory spinal roots on the surface of the body (dermatomes). (Reproduced with permission from D Sinclair: *Mechanisms of Cutaneous Sensation*. Oxford, UK, Oxford University Press, 1981 through PLS Clear.)

minimizing pressure on the skin. In general, it is better to avoid testing touch on hairy skin because of the profusion of the sensory endings that surround each hair follicle. The patient is tested with the eyes closed and should respond as soon as the stimulus is perceived, indicating its location.

Joint position testing is a measure of proprioception. With the patient's eyes closed, joint position is tested in the distal interphalangeal joint of the great toe and fingers. The digit is held by its sides, distal to the joint being tested, and moved passively while more proximal joints are stabilized—the patient indicates the change in position or direction of movement. If errors are made, more proximal joints are tested. A test of proximal joint position sense, primarily at the shoulder, is performed by asking the patient to bring the two index fingers together with arms extended and eyes closed. Normal individuals can do this accurately, with errors of 1 cm or less.

The sense of vibration is tested with an oscillating tuning fork that vibrates at 128 Hz. Vibration is tested over bony points, beginning distally; in the feet, it is tested over the dorsal surface of the distal phalanx of the big toes and at the malleoli of the ankles, and in the hands, it is tested dorsally at the distal phalanx of the fingers. If abnormalities are found, more proximal sites should be examined. Vibratory thresholds at the same site in the patient and the examiner may be compared for control purposes.

CORTICAL SENSATION

The most commonly used tests of cortical function are two-point discrimination, touch localization, and bilateral simultaneous stimulation, and tests for graphesthesia and stereognosis. Abnormalities

of these sensory tests, in the presence of normal primary sensation in an alert cooperative patient, signify a lesion of the parietal cortex or thalamocortical projections. If primary sensation is altered, these cortical discriminative functions usually will be abnormal also. Comparisons should always be made between analogous sites on the two sides of the body because the deficit with a specific parietal lesion is likely to be unilateral.

Two-point discrimination can be tested with calipers, the points of which may be set from 2 mm to several centimeters apart and then applied simultaneously to the test site. On the fingertips, a normal individual can distinguish about a 3-mm separation of points.

Touch localization is performed by light pressure for an instant with the examiner's fingertip or a wisp of cotton wool; the patient, whose eyes are closed, is required to identify the site of touch. *Bilateral simultaneous stimulation* at analogous sites (e.g., the dorsum of both hands) can be carried out to determine whether the perception of touch is extinguished consistently on one side (*extinction* or *neglect*). *Graphesthesia* refers to the capacity to recognize, with eyes closed, letters or numbers drawn by the examiner's fingertip on the palm of the hand. Once again, interside comparison is of prime importance. Inability to recognize numbers or letters is termed *agraphesthesia*.

Stereognosis refers to the ability to identify common objects by palpation, recognizing their shape, texture, and size. Common standard objects such as keys, paper clips, and coins are best used. Patients with normal stereognosis should be able to distinguish a dime from a penny and a nickel from a quarter without looking. Patients should feel the object with only one hand at a time. If they are unable to identify it in one hand, it should be placed in the other for comparison. Individuals who are unable to identify common objects and coins in one hand but can do so in the other are said to have *astereognosis* of the abnormal hand.

QUANTITATIVE SENSORY TESTING

Effective sensory testing devices are commercially available. Quantitative sensory testing is particularly useful for serial evaluation of cutaneous sensation in clinical trials. Threshold testing for touch and vibratory and thermal sensation is the most widely used application.

ELECTRODIAGNOSTIC STUDIES AND NERVE BIOPSY

Nerve conduction studies and nerve biopsy are important means of investigating the peripheral nervous system, but they do not evaluate the function or structure of cutaneous receptors and free nerve endings or of unmyelinated or thinly myelinated nerve fibers in the nerve trunks. Skin biopsy can be used to evaluate these structures in the dermis and epidermis.

LOCALIZATION OF SENSORY ABNORMALITIES

Sensory symptoms and signs can result from lesions at many different levels of the nervous system from the parietal cortex to the peripheral sensory receptor. Noting their distribution and nature is the most important way to localize their source. Their extent, configuration, symmetry, quality, and severity are the key observations.

Dysesthesias without sensory findings by examination may be difficult to interpret. To illustrate, tingling dysesthesias in an acral distribution (hands and feet) can be systemic in origin, for example, secondary to hyperventilation, or induced by a medication such as acetazolamide. Distal dysesthesias can also be an early event in an evolving polyneuropathy or may herald a myopathy, such as from vitamin B₁₂ deficiency. Sometimes, distal dysesthesias have no definable basis. In contrast, dysesthesias that correspond in distribution to that of a particular peripheral nerve structure denote a lesion at that site. For instance, dysesthesias restricted to the fifth digit and the adjacent one-half of the fourth finger on one hand reliably point to disorder of the ulnar nerve, most commonly at the elbow.

Nerve and Root In focal nerve trunk lesions, sensory abnormalities are readily mapped and generally have discrete boundaries

(Figs. 25-2 and 25-3). Root ("radicular") lesions frequently are accompanied by deep, aching pain along the course of the related nerve trunk. With compression of a fifth lumbar (L5) or first sacral (S1) root, as from a ruptured intervertebral disk, sciatica (radicular pain relating to the sciatic nerve trunk) is a common manifestation (Chap. 17). With a lesion affecting a single root, sensory deficits may be minimal or absent because adjacent root territories overlap extensively.

Isolated mononeuropathies may cause symptoms beyond the territory supplied by the affected nerve, but abnormalities on examination typically are confined to expected anatomic boundaries. In multiple mononeuropathies, symptoms and signs occur in discrete territories supplied by different individual nerves and—as more nerves are affected—may simulate a polyneuropathy if deficits become confluent. With polyneuropathies, sensory deficits are generally graded, distal, and symmetric in distribution (Chap. 446). Dysesthesias, followed by numbness, begin in the toes and ascend symmetrically. When dysesthesias reach the knees, they usually also have appeared in the fingertips. The process is nerve length-dependent, and the deficit is often described as "stocking glove" in type. Involvement of both hands and feet also occurs with lesions of the upper cervical cord or the brainstem, but an upper level of the sensory disturbance may then be found on the trunk and other evidence of a central lesion may be present, such as sphincter involvement or signs of an upper motor neuron lesion (Chap. 24). Although most polyneuropathies are pansensory and affect all modalities of sensation, selective sensory dysfunction according to nerve fiber size may occur. Small-fiber polyneuropathies are characterized by burning, painful dysesthesias with reduced pinprick and thermal sensation but with sparing of proprioception, motor function, and deep tendon reflexes. Touch is involved variably; when it is spared, the sensory pattern is referred to as exhibiting *sensory dissociation*. Sensory dissociation may occur also with spinal cord lesions (Chap. 442). Large-fiber polyneuropathies are characterized by vibration and position sense deficits, imbalance, absent tendon reflexes, and variable motor dysfunction but preservation of most cutaneous sensation. Dysesthesias, if present at all, tend to be tingling or bandlike in quality.

Sensory neuronopathy (or ganglionopathy) is characterized by widespread but asymmetric sensory loss occurring in a non-length-dependent manner so that it may occur proximally or distally, and in the arms, legs, or both. Pain and numbness progress to sensory ataxia and impairment of all sensory modalities over time. This condition is usually paraneoplastic or idiopathic in origin (Chaps. 94 and 445) or related to an autoimmune disease, particularly Sjögren's syndrome (Chap. 361).

Spinal Cord (See also Chap. 442) If the spinal cord is transected, all sensation is lost below the level of transection. Bladder and bowel function also are lost, as is motor function. Lateral hemisection of the spinal cord produces the Brown-Séquard syndrome, with absent pain and temperature sensation contralaterally and loss of proprioceptive sensation and power ipsilaterally below the lesion (see Figs. 25-1 and 442-1); ipsilateral pain or hyperesthesia may also occur.

Numbness or paresthesias in both feet may arise from a spinal cord lesion; this is especially likely when the upper level of the sensory loss extends to the trunk. When all extremities are affected, the lesion is probably in the cervical region or brainstem unless a peripheral neuropathy is responsible. The presence of upper motor neuron signs (Chap. 24) supports a central lesion; a hyperesthetic band on the trunk may suggest the level of involvement.

A dissociated sensory loss can reflect spinothalamic tract involvement in the spinal cord, especially if the deficit is unilateral and has an upper level on the torso. Bilateral spinothalamic tract involvement occurs with lesions affecting the center of the spinal cord, such as in syringomyelia. There is a dissociated sensory loss with impairment of pinprick and temperature appreciation but relative preservation of light touch, position sense, and vibration appreciation.

Dysfunction of the posterior columns in the spinal cord or of the posterior root entry zone may lead to a bandlike sensation around the trunk or a feeling of tight pressure in one or more limbs. Flexion

of the neck sometimes leads to an electric shock-like sensation that radiates down the back and into the legs (Lhermitte's sign) in patients with a cervical lesion affecting the posterior columns, such as from multiple sclerosis, cervical spondylosis, or following irradiation to the cervical region.

Brainstem Crossed patterns of sensory disturbance, in which one side of the face and the opposite side of the body are affected, localize to the lateral medulla. Here a small lesion may damage both the ipsilateral descending trigeminal tract and the ascending spinothalamic fibers subserving the opposite arm, leg, and hemitorso (see "Lateral medullary syndrome" in Fig. 426-7). A lesion in the tegmentum of the pons and midbrain, where the lemniscal and spinothalamic tracts merge, causes pansensory loss contralaterally.

Thalamus Hemisensory disturbance with tingling numbness from head to foot is often thalamic in origin but also can arise from the anterior parietal region. If abrupt in onset, the lesion is likely to be due to a small stroke (lacunar infarction), particularly if localized to the thalamus. Occasionally, with lesions affecting the VPL nucleus or adjacent white matter, a syndrome of thalamic pain, also called *Déjerine-Roussy syndrome*, may ensue. The persistent, unrelenting unilateral pain often is described in dramatic terms.

Cortex With lesions of the parietal lobe involving either the cortex or subjacent white matter, the most prominent symptoms are contralateral hemineglect, hemi-inattention, and a tendency not to use the affected hand and arm. On cortical sensory testing (e.g., two-point discrimination, graphesthesia), abnormalities are often found but primary sensation is usually intact. Anterior parietal infarction may present as a pseudothalamic syndrome with contralateral loss of primary sensation from head to toe. Dysesthesias or a sense of numbness and, rarely, a painful state may also occur.

Focal Sensory Seizures These seizures generally are due to lesions in the area of the postcentral or precentral gyrus. The principal symptom of focal sensory seizures is tingling, but additional, more complex sensations may occur, such as a rushing feeling, a sense of warmth, or a sense of movement without detectable motion. Symptoms typically are unilateral; commonly begin in the arm or hand, face, or foot; and often spread in a manner that reflects the cortical representation of different bodily parts, as in a Jacksonian march. Their duration is variable; seizures may be transient, lasting only for seconds, or persist for an hour or more. Focal motor features may supervene, often becoming generalized with loss of consciousness and tonic-clonic jerking.

Psychogenic Symptoms Sensory symptoms may have a psychogenic basis. Such symptoms may be generalized or have an anatomic boundary that is difficult to explain neurologically, for example, circumferentially at the groin or shoulder or around a specific joint. Pain is common, but the nature and intensity of any sensory disturbances are variable. The diagnosis should not be one of exclusion but based on suggestive findings that are otherwise difficult to explain, such as midline splitting of impaired vibration, pinprick, or light touch appreciation; variability or poor reproducibility of sensory deficits; or normal performance of tasks requiring sensory input that is seemingly abnormal on formal testing, such as good performance with eyes closed of the finger-to-nose test despite an apparent loss of position sense in the upper limb. The side with abnormal sensation may be confused when the limbs are placed in an unusual position, such as crossed behind the back. Sensory complaints should not be regarded as psychogenic simply because they are unusual.

TREATMENT

Management is based on treatment of the underlying condition. Symptomatic treatment of acute and chronic pain is discussed in Chap. 13. Dysesthesias, when severe and persistent, may respond to anticonvulsants (carbamazepine, 100–1000 mg/d; gabapentin, 300–3600 mg/d; or pregabalin, 50–300 mg/d), antidepressants (amitriptyline, 25–150 mg/d; nortriptyline, 25–150 mg/d; desipramine, 100–300 mg/d; or venlafaxine, 75–225 mg/d).

ACKNOWLEDGMENTS

The editors acknowledge the contributions of Michael J. Aminoff to earlier editions of this chapter.

FURTHER READING

- BRAZIS P et al: *Localization in Clinical Neurology*, 7th ed. Philadelphia, Lippincott William & Wilkins, 2016.
- CAMPBELL WW, BAROHN RJ: *DeJong's the Neurologic Examination*, 8th ed. Philadelphia, Wolters Kluwer, 2020.
- WAXMAN S: *Clinical Neuroanatomy*, 29th ed. New York, McGraw Hill Education, 2020.

26

Gait Disorders, Imbalance, and Falls

Jessica M. Baker



PREVALENCE, MORBIDITY, AND MORTALITY

Gait and balance problems are common in the elderly and contribute to the risk of falls and injury. Gait disorders have been described in 15% of individuals aged >65. By age 80, one person in four will use a mechanical aid to assist with ambulation. Among those aged ≥85, the prevalence of gait abnormality approaches 40%. In epidemiologic studies, gait disorders are consistently identified as a major risk factor for falls and injury.

ANATOMY AND PHYSIOLOGY

An upright bipedal gait depends on the successful integration of postural control and locomotion. These functions are widely distributed in the central nervous system. The biomechanics of bipedal walking are complex, and the performance is easily compromised by a neurologic deficit at any level. Command and control centers in the brainstem, cerebellum, and forebrain modify the action of spinal pattern generators to promote stepping. While a form of "fictive locomotion" can be elicited from quadrupedal animals after spinal transection, this capacity is limited in primates. Step generation in primates is dependent on locomotor centers in the pontine tegmentum, midbrain, and subthalamic region. Locomotor synergies are executed through the reticular formation and descending pathways in the ventromedial spinal cord. Cerebral control provides a goal and purpose for walking and is involved in avoidance of obstacles and adaptation of locomotor programs to context and terrain.

Postural control requires the maintenance of the center of mass over the base of support through the gait cycle. Unconscious postural adjustments maintain standing balance: long latency responses are measurable in the leg muscles, beginning 110 milliseconds after a perturbation. Forward motion of the center of mass provides propulsive force for stepping, but failure to maintain the center of mass within stability limits results in falls. The anatomic substrate for dynamic balance has not been well defined, but the vestibular nucleus and midline cerebellum contribute to balance control in animals. Patients with damage to these structures have impaired balance while standing and walking.

Standing balance depends on good-quality sensory information about the position of the body center with respect to the environment, support surface, and gravitational forces. Sensory information for postural control is primarily generated by the visual system, the vestibular system, and proprioceptive receptors in the muscle spindles and joints. A healthy redundancy of sensory afferent information is generally available, but loss of two of the three pathways is sufficient to compromise standing balance. Balance disorders in older individuals

sometimes result from multiple insults in the peripheral sensory systems (e.g., visual loss, vestibular deficit, peripheral neuropathy) that critically degrade the quality of afferent information needed for balance stability.

Older patients with cognitive impairment appear to be particularly prone to falls and injury. There is a growing body of literature on the use of attentional resources to manage gait and balance. Walking is generally considered to be unconscious and automatic, but the ability to walk while attending to a cognitive task (*dual-task walking*) may be compromised in the elderly. Older patients with deficits in executive function may have particular difficulty in managing the attentional resources needed for dynamic balance when distracted.

DISORDERS OF GAIT

Disorders of gait may be attributed to neurologic and nonneurologic causes, although significant overlap often exists. The *antalgic gait* results from avoidance of pain associated with weight bearing and is commonly seen in osteoarthritis. Asymmetry is a common feature of gait disorders due to contractures and other orthopedic deformities. Impaired vision rounds out the list of common nonneurologic causes of gait disorders.

Neurologic gait disorders are disabling and equally important to address. The heterogeneity of gait disorders observed in clinical practice reflects the large network of neural systems involved in the task. Walking is vulnerable to neurologic disease at every level. Gait disorders have been classified descriptively on the basis of abnormal physiology and biomechanics. One problem with this approach is that many failing gaits look fundamentally similar. This overlap reflects common patterns of adaptation to threatened balance stability and declining performance. *The gait disorder observed clinically must be viewed as the product of a neurologic deficit and a functional adaptation.* Unique features of the failing gait are often overwhelmed by the adaptive response. Some common patterns of abnormal gait are summarized next. Gait disorders can also be classified by etiology (Table 26-1).

CAUTIOUS GAIT

The term *cautious gait* is used to describe the patient who walks with an abbreviated stride, widened base, and lowered center of mass, as if walking on a slippery surface. Arms are often held abducted. This disorder is both common and nonspecific. It is, in essence, an adaptation to a perceived postural threat. There may be an associated fear of falling. This disorder can be observed in more than one-third of older

patients with gait impairment. Physical therapy often improves walking to the degree that follow-up observation may reveal a more specific underlying disorder.

STIFF-LEGGED GAIT

Spastic gait is characterized by stiffness in the legs, an imbalance of muscle tone, and a tendency to circumduct and scuff the feet. The disorder reflects compromise of corticospinal command and overactivity of spinal reflexes. The patient may walk on the toes. In extreme instances, the legs cross due to increased tone in the adductors (“scissoring” gait). Upper motor neuron signs are present on physical examination. The disorder may be cerebral or spinal in origin.

Myelopathy from cervical spondylosis is a common cause of spastic or spastic-ataxic gait in the elderly. Demyelinating disease and trauma are the leading causes of myelopathy in younger patients. In chronic progressive myelopathy of unknown cause, a workup with laboratory and imaging tests may establish a diagnosis. A structural lesion, such as a tumor or a spinal vascular malformation, should be excluded with appropriate testing. **Spinal cord disorders are discussed in detail in Chap. 442.**

With cerebral spasticity, asymmetry is common, the upper extremities are usually involved, and dysarthria is often an associated feature. Common causes include vascular disease (stroke), multiple sclerosis, motor neuron disease, and perinatal nervous system injury (cerebral palsy).

Other stiff-legged gaits include dystonia (Chap. 436) and stiff-person syndrome (Chap. 94). Dystonia is a disorder characterized by sustained muscle contractions resulting in repetitive twisting movements and abnormal posture. It often has a genetic basis. Dystonic spasms can produce plantar flexion and inversion of the feet, sometimes with torsion of the trunk. In autoimmune stiff-person syndrome, exaggerated lordosis of the lumbar spine and overactivation of antagonist muscles restrict trunk and lower-limb movement and result in a wooden or fixed posture.

PARKINSONISM, FREEZING GAIT, AND OTHER MOVEMENT DISORDERS

Parkinson's disease (Chap. 435) is common, affecting 1% of the population >65 years of age. The stooped posture, shuffling gait, and decreased arm swing are characteristic and distinctive features. Patients sometimes accelerate (festinate) with walking, display retropulsion, or exhibit a tendency to turn en bloc. The step-to-step variability

TABLE 26-1 Prevalence of Neurologic Gait Disorders

NEUROLOGIC GAIT DISORDER	NO. (%) ^a	TOTAL NUMBER ^b	CAUSES (NO.)
Single neurologic gait disorder	81 (69%)		
Sensory ataxic	22 (18%)	46	Peripheral sensory neuropathy (46)
Parkinsonian	19 (16%)	34	Parkinson's disease (18), drug-induced parkinsonism (8), dementia with parkinsonism (4), parkinsonism (4)
Higher level	9 (8%)	31	Vascular encephalopathy (20), normal pressure hydrocephalus (1), severe dementia (7), hypoxic ischemic encephalopathy (1), unknown (1)
Cerebellar ataxic	7 (6%)	10	Cerebellar stroke (3), cerebellar lesion due to multiple sclerosis (1), severe essential tremor (3), postvaccinal cerebellitis (1), chronic alcohol abuse (1), multiple system atrophy (1)
Cautious	7 (6%)	7	Idiopathic, associated fear of falling (7)
Paretic/hypotonic	6 (5%)	14	Neurogenic claudication (7), diabetic neuropathy (1), nerve lesion due to trauma or surgery (4), distal paraparesis after Guillain-Barré syndrome (1), unknown (2)
Spastic	6 (5%)	7	Ischemic stroke (3), intracerebral hemorrhage (3), congenital (1)
Vestibular ataxic	4 (3%)	6	Bilateral vestibulopathy (3), recent vestibular neuritis (1), recent Ménière's attack (1), acoustic neuroma with surgery (1)
Dyskinetic	1 (1%)	4	Levodopa-induced dyskinesia (3), chorea (1)
Multiple neurologic gait disorders	36 (30%)		
Total	117		

^aPercentage of individuals with a single gait disorder. ^bIncludes individuals with multiple gait disorders.

Note: Of 117 patients with a neurologic gait disorder, 81 had a single neurologic gait disorder; the remainder (36) had multiple neurologic gait disorders.

Source: Reproduced with modifications from P Mahlknecht et al: PLoS One 8:e69627, 2013.

of the parkinsonian gait also contributes to falls, which are a major source of morbidity, particularly later in the disease course. Dopamine replacement improves step length, arm swing, turning speed, and gait initiation. There is increasing evidence that deficits in cholinergic circuits in the pedunculopontine nucleus and cortex contribute to the gait disorder of Parkinson's disease. Cholinesterase inhibitors such as donepezil and rivastigmine have been shown in early studies to significantly decrease gait variability, instability, and fall frequency, even in the absence of cognitive impairment, perhaps through improvement in attention.

Freezing is defined as a brief, episodic absence of forward progression of the feet, despite the intention to walk. Freezing may be triggered by approaching a narrow doorway or crowd, may be overcome by visual cueing, and contributes to fall risk. Gait freezing is present in approximately one-quarter of Parkinson's patients within 5 years of onset, and its frequency increases further over time. In treated patients, end-of-dose gait freezing is a common problem that may improve with more frequent administration of dopaminergic drugs or with use of monoamine oxidase type B inhibitors such as rasagiline or selegiline (Chap. 435).

Freezing of gait is also common in other neurodegenerative disorders associated with parkinsonism, including progressive supranuclear palsy (PSP), multiple-system atrophy, and corticobasal degeneration. Patients with these disorders frequently present with axial stiffness, postural instability, and a shuffling, freezing gait while lacking the characteristic pill-rolling tremor of Parkinson's disease. The gait of PSP is typically more erect compared with the stooped posture of typical Parkinson's disease, and falls within the first year also suggest the possibility of PSP. The gait of vascular parkinsonism tends to be broad-based and shuffling with reduced arm swing bilaterally; disproportionate involvement of gait early in the disease course differentiates this entity from Parkinson's disease.

Hyperkinetic movement disorders also produce characteristic and recognizable disturbances in gait. In Huntington's disease (Chap. 436), the unpredictable occurrence of choreic movements gives the gait a dancing quality. Tardive dyskinesia is the cause of many odd, stereotypic gait disorders seen in patients chronically exposed to antipsychotics and other drugs that block the D₂ dopamine receptor. *Orthostatic tremor* is a high-frequency, low-amplitude tremor predominantly involving the lower extremities. Patients often report shakiness or unsteadiness on standing and improvement with sitting or walking. Falls are common. The tremor is often only appreciable by palpating the legs while standing.

■ FRONTAL GAIT DISORDER

Frontal gait disorder, also known as higher-level gait disorder, is common in the elderly and has a variety of causes. The term is used to describe a shuffling, freezing gait with imbalance, and other signs of higher cerebral dysfunction. Typical features include a wide base of support, a short stride, shuffling along the floor, and difficulty with starts and turns. Many patients exhibit a difficulty with gait initiation that is descriptively characterized as the "slipping clutch" syndrome or gait ignition failure. The term *lower-body parkinsonism* is also used to describe such patients. Strength is generally preserved, and patients are able to

make stepping movements when not standing and maintaining their balance at the same time. This disorder is best considered a higher-level motor control disorder, as opposed to an apraxia (Chap. 30), though the term *gait apraxia* persists in the literature.

The most common cause of frontal gait disorder is vascular disease, particularly subcortical small-vessel disease in the deep frontal white matter and centrum ovale. Over three-quarters of patients with subcortical vascular dementia demonstrate gait abnormalities; decreased arm swing and a stooped posture are particularly prevalent features. The clinical syndrome also includes dysarthria, pseudobulbar affect (emotional disinhibition), increased tone, and hyperreflexia in the lower limbs.

Normal pressure (communicating) hydrocephalus (NPH) in adults also presents with a similar gait disorder (Chap. 431). Other features of the diagnostic triad (mental changes, incontinence) may be absent in a substantial number of patients. MRI demonstrates ventricular enlargement, an enlarged flow void about the aqueduct, periventricular white matter change, and high-convexity tightness (disproportionate widening of the sylvian fissures versus the cortical sulci). A lumbar puncture or dynamic test is necessary to confirm a diagnosis of NPH. Neurodegenerative dementias and mass lesions of the frontal lobes cause a similar clinical picture and can be differentiated from vascular disease and hydrocephalus by neuroimaging.

■ CEREBELLAR GAIT ATAXIA

Disorders of the cerebellum (Chap. 439) have a dramatic impact on gait and balance. Cerebellar gait ataxia is characterized by a wide base of support, lateral instability of the trunk, erratic foot placement, and decompensation of balance when attempting to walk on a narrow base. Difficulty maintaining balance when turning is often an early feature. Patients are unable to walk tandem heel to toe and display truncal sway in narrow-based or tandem stance. They show considerable variation in their tendency to fall in daily life.

Causes of cerebellar ataxia in older patients include stroke, trauma, tumor, and neurodegenerative disease such as multiple-system atrophy (Chap. 440) and various forms of hereditary cerebellar degeneration (Chap. 439). A short expansion at the site of the fragile X mutation (*fragile X premutation*) has been associated with gait ataxia in older men. Alcohol causes an acute and chronic cerebellar ataxia. In patients with ataxia due to cerebellar degeneration, MRI demonstrates the extent and topography of cerebellar atrophy.

■ SENSORY ATAXIA

As reviewed earlier in this chapter, balance depends on high-quality afferent information from the visual and the vestibular systems and proprioception. When this information is lost or degraded, balance during locomotion is impaired and instability results. The sensory ataxia of tabetic neurosyphilis is a classic example. The contemporary equivalent is the patient with neuropathy affecting large fibers. Vitamin B₁₂ deficiency is a treatable cause of large-fiber sensory loss in the spinal cord and peripheral nervous system. Joint position and vibration sense are diminished in the lower limbs. The stance in such patients is destabilized by eye closure; they often look down at their feet when walking and do poorly in the dark. Table 26-2 compares sensory ataxia with cerebellar ataxia and frontal gait disorder.

TABLE 26-2 Features of Cerebellar Ataxia, Sensory Ataxia, and Frontal Gait Disorders

FEATURE	CEREBELLAR ATAXIA	SENSORY ATAXIA	FRONTAL GAIT
Base of support	Wide-based	Narrow base, looks down	Wide-based
Velocity	Variable	Slow	Very slow
Stride	Irregular, lurching	Regular with path deviation	Short, shuffling
Romberg test	+/-	Unsteady, falls	+/-
Heel → shin	Abnormal	+/-	Normal
Initiation	Normal	Normal	Hesitant
Turns	Unsteady	+/-	Hesitant, multistep
Postural instability	+	+++	++++ Poor postural synergies rising from a chair
Falls	Late event	Frequent	Frequent

■ NEUROMUSCULAR DISEASE

Patients with neuromuscular disease often have an abnormal gait, occasionally as a presenting feature. With distal weakness (peripheral neuropathy), the step height is increased to compensate for foot drop, and the sole of the foot may slap on the floor during weight acceptance, termed the *steppage gait*. Patients with myopathy or muscular dystrophy more typically exhibit proximal weakness. Weakness of the hip girdle may result in some degree of excess pelvic sway during locomotion. The stooped posture of lumbar spinal stenosis ameliorates pain from the compression of the cauda equina occurring with a more upright posture while walking and may mimic early parkinsonism.

■ TOXIC AND METABOLIC DISORDERS

Chronic toxicity from medications and metabolic disturbances can impair motor function and gait. Examination may reveal mental status changes, asterixis, or myoclonus. Static equilibrium is disturbed, and such patients are easily thrown off balance. Disequilibrium is particularly evident in patients with chronic renal disease and those with hepatic failure, in whom asterixis may impair postural support. Sedative drugs, especially neuroleptics and long-acting benzodiazepines, affect postural control and increase the risk for falls. These disorders are especially important to recognize because they are often treatable.

■ FUNCTIONAL GAIT DISORDER

Functional neurologic disorders (formerly “psychogenic”) are common in practice, and the presentation often involves gait. Sudden onset, inconsistent deficits, waxing and waning course, incongruence of symptoms with an organic lesion, and improvement with distraction are key features. Phenomenology is variable; extreme slow motion, an inappropriately overcautious gait, odd gyrations of posture with wastage of muscular energy, astasia-abasia (inability to stand and walk), bouncing, and foot stiffness (dystonia) have been described. Falls are rare, and there are often discrepancies between examination findings and the patient’s functional status. Preceding stress or trauma is variably present, and its absence does not preclude the diagnosis of a functional gait disorder. Functional gait disorders may be challenging to diagnose and should be differentiated from the slowness and psychomotor retardation seen in certain patients with major depression.

APPROACH TO THE PATIENT

Slowly Progressive Disorder of Gait

When reviewing the history, it is helpful to inquire about the onset and progression of disability. Initial awareness of an unsteady gait often follows a fall. Stepwise evolution or sudden progression suggests vascular disease. Gait disorder may be associated with urinary urgency and incontinence, particularly in patients with cervical spine disease or hydrocephalus. It is always important to review the use of alcohol and medications that affect gait and balance. Information on localization derived from the neurologic examination can be helpful in narrowing the list of possible diagnoses.

Gait observation provides an immediate sense of the patient’s degree of disability. Arthritic and antalgic gaits are recognized by observation, although neurologic and orthopedic problems may coexist. Characteristic patterns of abnormality are sometimes seen, although, as stated previously, failing gaits often look fundamentally similar. Cadence (steps per minute), velocity, and stride length can be recorded by timing a patient over a fixed distance. Watching the patient rise from a chair provides a good functional assessment of balance.

Brain imaging studies may be informative in patients with an undiagnosed disorder of gait. MRI is sensitive for cerebral lesions of vascular or demyelinating disease and is a good screening test for occult hydrocephalus. Patients with recurrent falls are at risk for subdural hematoma. As mentioned earlier, many elderly patients with gait and balance difficulty have white matter abnormalities in the periventricular region and centrum semiovale. While these lesions may be an incidental finding, a substantial burden of white matter disease will ultimately impact cerebral control of locomotion.

DISORDERS OF BALANCE

■ DEFINITION, ETIOLOGY, AND MANIFESTATIONS

Balance is the ability to maintain equilibrium—a dynamic state in which one’s center of mass is controlled with respect to the lower extremities, gravity, and the support surface despite external perturbations. The reflexes required to maintain upright posture require input from cerebellar, vestibular, and somatosensory systems; the premotor cortex and corticospinal and reticulospinal tracts mediate output to axial and proximal limb muscles. These responses are physiologically complex, and the anatomic representation they entail is not well understood. Failure can occur at any level and presents as difficulty maintaining posture while standing and walking.

The history and physical examination may differentiate underlying causes of imbalance. Patients with *cerebellar* ataxia do not generally complain of dizziness, although balance is visibly impaired. Neurologic examination reveals a variety of cerebellar signs. Postural compensation may prevent falls early on, but falls are inevitable with disease progression. The progression of neurodegenerative ataxia is often measured by the number of years to loss of stable ambulation.

Vestibular disorders (Chap. 22) have symptoms and signs that fall into three categories: (1) vertigo (the subjective inappropriate perception or illusion of movement); (2) nystagmus (involuntary eye movements); and (3) impaired standing balance. Not every patient has all manifestations. Patients with vestibular deficits related to ototoxic drugs may lack vertigo or obvious nystagmus, but their balance is impaired on standing and walking, and they cannot navigate in the dark. Laboratory testing is available to investigate vestibular deficits.

Somatosensory deficits also produce imbalance and falls. There is often a subjective sense of insecure balance and fear of falling. Postural control is compromised by eye closure (*Romberg’s sign*); these patients also have difficulty navigating in the dark. A dramatic example is provided by the patient with autoimmune subacute sensory neuropathy, which is sometimes a paraneoplastic disorder (Chap. 94). Compensatory strategies enable such patients to walk in the virtual absence of proprioception, but the task requires active visual monitoring.

Patients with *higher-level disorders of equilibrium* have difficulty maintaining balance in daily life and may present with falls. Their awareness of balance impairment may be reduced. Patients taking sedating medications are in this category.

■ FALLS

Falls are common in the elderly; over one-third of people aged >65 who are living in the community fall each year. This number is even higher in nursing homes and hospitals. Elderly people are not only at higher risk for falls but are also more likely to suffer serious complications due to medical comorbidities such as osteoporosis. Hip fractures result in hospitalization, can lead to nursing home admission, and are associated with an increased mortality risk in the subsequent year. Falls may result in brain or spinal injury, the history of which may be difficult for the patient to provide. The proportion of spinal cord injuries due to falls in individuals aged >65 years has doubled in the past decade, perhaps due to increasing activity in this age group. Some falls result in a prolonged time lying on the ground; fractures and CNS injury are a particular concern in this context.

For each person who is physically disabled, there are others whose functional independence is limited by anxiety and fear of falling. Nearly one in five elderly individuals voluntarily restricts his or her activity because of fear of falling. With loss of ambulation, the quality of life diminishes, and rates of morbidity and mortality increase.

■ RISK FACTORS FOR FALLS

Risk factors for falls may be *intrinsic* (e.g., gait and balance disorders) or *extrinsic* (e.g., polypharmacy, environmental factors); some risk factors are modifiable. The presence of multiple risk factors is associated with a substantially increased risk of falls. Table 26-3 summarizes a meta-analysis of studies establishing the principal risk factors for falls. Polypharmacy (use of four or more prescription medications) has also been identified as an important risk factor.

TABLE 26-3 Meta-Analysis of Risk Factors for Falls in Older Persons

RISK FACTOR	MEAN RR (OR)	RANGE
Muscle weakness	4.4	1.5–10.3
History of falls	3.0	1.7–7.0
Gait deficit	2.9	1.3–5.6
Balance deficit	2.9	1.6–5.4
Use assistive device	2.6	1.2–4.6
Visual deficit	2.5	1.6–3.5
Arthritis	2.4	1.9–2.9
Impaired ADL	2.3	1.5–3.1
Depression	2.2	1.7–2.5
Cognitive impairment	1.8	1.0–2.3
Age >80 years	1.7	1.1–2.5

Abbreviations: ADL, activity of daily living; OR, odds ratio from retrospective studies; RR, relative risk from prospective studies.

Source: Reproduced with permission from Guideline for the Prevention of Falls in Older Persons. J Am Geriatr Soc 49:664, 2001.

ASSESSMENT OF THE PATIENT WITH FALLS

The most productive approach is to identify the high-risk patient prospectively, before there is a serious injury. All community-dwelling adults should be asked annually about falls and whether or not fear of falling limits daily activities. The Timed Up and Go ("TUG") test involves timing a patient as they stand up from a chair, walk 10 feet, turn, and then sit down. Patients with a history of falls or those requiring >12 s to complete the TUG test are at high risk for falls and should undergo further assessment.

History The history surrounding a fall is often problematic or incomplete, and the underlying mechanism or cause may be difficult to establish in retrospect. Patients should be queried about any provoking factors (including head turn, standing) or prodromal symptoms, such as dizziness, vertigo, presyncopal symptoms, or focal weakness. A history of baseline mobility and medical comorbidities should be elicited. Patients at particular risk include those with mental status changes or dementia. Medications should be reviewed, with particular attention to benzodiazepines, opioids, antipsychotics, antiepileptics, antidepressants, antiarrhythmics, and diuretics, all of which are associated with an increased risk of falls. It is equally important to distinguish *mechanical falls* (those caused by tripping or slipping) due to purely extrinsic or environmental factors from those in which a modifiable intrinsic factor contributes. *Recurrent falls* may indicate an underlying gait or balance disorder. Falls associated with loss of consciousness (syncope, seizure) may require appropriate cardiac or neurologic evaluation and intervention (**Chaps. 21 and 425**), although a patient's report of change in consciousness may be unreliable.

Physical Examination Examination of the patient with falls should include a basic cardiac examination, including orthostatic blood pressure if indicated by history, and observation of any orthopedic abnormalities. Mental status is easily assessed while obtaining a history from the patient; the remainder of the neurologic examination should include visual acuity, strength and sensation in the lower extremities, muscle tone, and cerebellar function, with particular attention to gait and balance as described earlier in this chapter.

Fall Patterns The description of a fall event may provide further clues to the underlying etiology. While there is no standard nosology of falls, some common clinical patterns may emerge and provide a clue.

DROP ATTACKS AND COLLAPSING FALLS Drop attacks and collapsing falls are associated with a sudden loss of postural tone. Patients may report that their legs just "gave out" underneath them or that they "collapsed in a heap." Syncope or orthostatic hypotension may be a factor in some such falls. Neurologic causes are relatively rare but include atonic seizures, myoclonus, and intermittent obstruction of the foramen of Monro by a colloid cyst of the third ventricle causing acute obstructive hydrocephalus. An emotional trigger suggests cataplexy.

While collapsing falls are more common among older patients with vascular risk factors, drop attacks should not be confused with vertebrobasilar ischemic attacks.

TOPPLING FALLS Some patients maintain tone in antigravity muscles but fall over like a tree trunk, as if postural defenses had disengaged. Causes include cerebellar pathology and lesions of the vestibular system. There may be a consistent direction to such falls. Toppling falls are an early feature of progressive supranuclear palsy, and a late feature of Parkinson's disease, once postural instability has developed. Thalamic lesions causing truncal instability (*thalamic astasia*) may also contribute to this type of fall.

FALLS DUE TO GAIT FREEZING Freezing of gait is seen in Parkinson's disease and related disorders. The feet stick to the floor and the center of mass keeps moving, resulting in a disequilibrium from which the patient has difficulty recovering, resulting in a forward fall. Similarly, patients with Parkinson's disease and festinating gait may find their feet unable to keep up and may thus fall forward.

FALLS RELATED TO SENSORY LOSS Patients with somatosensory, visual, or vestibular deficits are prone to falls. These patients have particular difficulty dealing with poor illumination or walking on uneven ground. They often report subjective imbalance, apprehension, and fear of falling. These patients may be especially responsive to a rehabilitation-based intervention.

FALLS RELATED TO WEAKNESS Patients who lack strength in antigravity muscles have difficulty rising from a chair or maintaining their balance after a perturbation. These patients are often unable to get up after a fall and may have to remain on the floor for a prolonged period until help arrives. If due to deconditioning, this is often treatable. Resistance strength training can increase muscle mass and leg strength, even for people in their eighties and nineties.

TREATMENT

Interventions to Reduce the Risk of Falls and Injury

Efforts should be made to define the mechanism underlying falls in a given patient, as specific treatment may be possible once a diagnosis is established. Orthostatic changes in blood pressure and pulse should be recorded. Medications (including over-the-counter) should be reviewed, reevaluating benefits and burdens of medications that might increase fall risk. Treatment of cataracts and avoidance of multifocal lenses could be considered for patients whose falls result from vision impairment. A home visit to look for environmental hazards can be helpful. A variety of modifications may be recommended to improve safety, including improved lighting, installation of grab bars and nonslip surfaces, and use of adaptive equipment.

Home- and group-based exercise programs focusing on leg strength and balance, physical therapy, and use of assistive devices reduce fall risk in individuals with a history of falls or disorders of gait and balance. Rehabilitative interventions aim to improve muscle strength and balance stability and to make the patient more resistant to injury. High-intensity resistance strength training with weights and machines is useful to improve muscle mass, even in frail older patients. Improvements realized in posture and gait should translate to reduced risk of falls and injury. Sensory balance training is another approach to improving balance stability. Measurable gains can be made in a few weeks of training, and benefits can be maintained over 6 months by a 10- to 20-min home exercise program. This strategy is particularly successful in patients with vestibular and somatosensory balance disorders. The National Institute on Aging provides online examples of balance exercises for older adults. A Tai Chi exercise program has been demonstrated to reduce the risk of falls and injury in patients with Parkinson's disease. Cognitive training, including dual-task training, may improve mobility in older adults with cognitive impairment.

I am grateful to Dr. Lewis R. Sudarsky for his substantial contributions to earlier versions of this chapter.

FURTHER READING

- AMERICAN GERIATRICS SOCIETY, BRITISH GERIATRICS SOCIETY, AMERICAN ACADEMY OF ORTHOPEDIC SURGEONS PANEL ON FALLS PREVENTION: Guideline for the prevention of falls in older persons. *J Am Geriatr Soc* 49:664, 2001.
- GANZ D, LATHAM N: Prevention of falls in community-dwelling older adults. *N Engl J Med* 382:734, 2020.
- NATIONAL INSTITUTE ON AGING: EXERCISE AND PHYSICAL ACTIVITY. Available from <https://www.nia.nih.gov/health/exercise-physical-activity>. Accessed April 25, 2021.
- NUTT JG: Classification of gait and balance disorders. *Adv Neurol* 87:135, 2001.
- PIRKER W, KATZENSCHLAGER R: Gait disorders in adults and the elderly. *Wien Klin Wochenschr* 129:81, 2017.

reversible cognitive disturbance. Hyperactive patients are often easily recognized by their characteristic severe agitation, tremor, hallucinations, and autonomic instability. Patients who are quietly hypoactive are more often overlooked on the medical wards and in the ICU.

The reversibility of delirium is emphasized because many etiologies, such as infection and medication effects, can be treated easily. The long-term cognitive consequences of delirium remain an area of active research. Some episodes of delirium continue for weeks, months, or even years. The persistence of delirium in some patients and its high recurrence rate may be due to inadequate initial treatment of the underlying etiology. In other instances, delirium appears to cause permanent neuronal damage and long-term cognitive decline. Therefore, prevention strategies are important to implement. Even if an episode of delirium completely resolves, there may be lingering effects of the disorder; a patient's recall of events after delirium varies widely, ranging from complete amnesia to repeated reexperiencing of the frightening period of confusion, similar to what is seen in patients with posttraumatic stress disorder.

RISK FACTORS

An effective primary prevention strategy for delirium begins with identification of high-risk patients. Some hospital systems have initiated comprehensive delirium programs that screen most or all patients upon admission or before elective surgery; positive screens trigger a host of focused prevention measures. Multiple validated scoring systems have been developed as a screen for asymptomatic patients, many of which emphasize well-established risk factors for delirium.

The two most consistently identified risk factors are older age and baseline cognitive dysfunction. Individuals who are aged >65 or exhibit low scores on standardized tests of cognition develop delirium upon hospitalization at a rate approaching 50%. Whether age and baseline cognitive dysfunction are truly independent risk factors is uncertain. Other predisposing factors include sensory deprivation, such as preexisting hearing and visual impairment, as well as indices for poor overall health, including baseline immobility, malnutrition, and underlying medical or neurologic illness.

In-hospital risks for delirium include the use of bladder catheterization, physical restraints, sleep and sensory deprivation, and the addition of three or more new medications. Avoiding such risks remains a key component of delirium prevention as well as treatment. Surgical and anesthetic risk factors for the development of postoperative delirium include procedures such as those involving cardiopulmonary bypass, inadequate or excessive treatment of pain in the immediate postoperative period, and perhaps specific agents such as inhalational anesthetics.

The relationship between delirium and dementia ([Chap. 29](#)) is complicated by significant overlap between the two conditions, and it is not always simple to distinguish between them. Dementia and preexisting cognitive dysfunction serve as major risk factors for delirium, and at least two-thirds of cases of delirium occur in patients with coexisting underlying dementia. A form of dementia with parkinsonism, *dementia with Lewy bodies* ([Chap. 434](#)), is characterized by a fluctuating course, prominent visual hallucinations, parkinsonism, and an attentional deficit that clinically resembles hyperactive delirium; patients with this condition are particularly vulnerable to delirium. Delirium in the elderly often reflects an insult to a brain that is vulnerable due to an underlying neurodegenerative condition. Therefore, the development of delirium sometimes heralds the onset of a previously unrecognized brain disorder, and after the acute delirious episode has cleared, careful screening for an underlying condition should occur in the outpatient setting.

EPIDEMIOLOGY

Delirium is common, but its reported incidence has varied widely with the criteria used to define this disorder. Estimates of delirium in hospitalized patients range from 10% to >50%, with higher rates reported for elderly patients and patients undergoing hip surgery. Older patients in the ICU have especially high rates of delirium that approach 75%. The

27

Confusion and Delirium

S. Andrew Josephson, Bruce L. Miller



Confusion, a mental and behavioral state of reduced comprehension, coherence, and capacity to reason, is one of the most common problems encountered in medicine, accounting for a large number of emergency department visits, hospital admissions, and inpatient consultations. **Delirium**, a term used to describe an acute confusional state, remains a major cause of morbidity and mortality, costing billions of dollars yearly in health care costs in the United States alone. Despite increased efforts targeting awareness of this condition, delirium often goes unrecognized in the face of evidence that it is usually the cognitive manifestation of serious underlying medical or neurologic illness.

CLINICAL FEATURES OF DELIRIUM

A multitude of terms are used to describe patients with delirium, including *encephalopathy*, *acute brain failure*, *acute confusional state*, and *postoperative or intensive care unit (ICU) psychosis*. Delirium has many clinical manifestations, but it is defined as a relatively acute decline in cognition that fluctuates over hours or days. The hallmark of delirium is a deficit of attention, although all cognitive domains—including memory, executive function, visuospatial tasks, and language—are variably involved. Associated symptoms that may be present in some cases include altered sleep-wake cycles, perceptual disturbances such as hallucinations or delusions, affect changes, and autonomic findings that include heart rate and blood pressure instability.

Delirium is a clinical diagnosis that is made only at the bedside. Two subtypes have been described—hyperactive and hypoactive—based on differential psychomotor features. The cognitive syndrome associated with severe alcohol withdrawal (i.e., “*delirium tremens*”) remains the classic example of the hyperactive subtype, featuring prominent hallucinations, agitation, and hyperarousal, often accompanied by life-threatening autonomic instability. In striking contrast is the hypoactive subtype, exemplified by benzodiazepine intoxication, in which patients are withdrawn and quiet, with prominent apathy and psychomotor slowing.

This dichotomy between subtypes of delirium is a useful construct, but patients often fall somewhere along a spectrum between the hyperactive and hypoactive extremes, sometimes fluctuating from one to the other. Therefore, clinicians must recognize this broad range of presentations of delirium to identify all patients with this potentially

condition is not recognized in up to one-third of delirious inpatients, and the diagnosis is especially problematic in the ICU environment, where cognitive dysfunction is often difficult to appreciate in the setting of serious systemic illness and sedation. Delirium in the ICU should be viewed as an important manifestation of organ dysfunction not unlike liver, kidney, or heart failure. Outside the acute hospital setting, delirium occurs in nearly one-quarter of patients in nursing homes and in 50–80% of those at the end of life. These estimates emphasize the remarkably high frequency of this cognitive syndrome in older patients, a population that continues to grow.

An episode of delirium was previously viewed as a transient condition that carried a benign prognosis. It is now recognized as a disorder with substantial morbidity and mortality, and that often represents the first manifestation of a serious underlying illness. Estimates of in-hospital mortality rates among delirious patients range from 25% to 33%, similar to mortality rates due to sepsis. Patients with an in-hospital episode of delirium have a fivefold higher mortality rate in the months after their illness compared with age matched nondelirious hospitalized patients. Delirious hospitalized patients also have a longer length of stay, are more likely to be discharged to a nursing home, have a higher frequency of readmission, and are more likely to experience subsequent episodes of delirium and cognitive decline; as a result, this condition has an enormous economic cost.

PATHOGENESIS

The pathogenesis and anatomy of delirium are incompletely understood. The attentional deficit that serves as the neuropsychological hallmark of delirium has a diffuse localization within the brainstem, thalamus, prefrontal cortex, and parietal lobes. Rarely, focal lesions such as ischemic strokes have led to delirium in otherwise healthy persons; right parietal and medial dorsal thalamic lesions have been reported most commonly, pointing to the importance of these areas in delirium pathogenesis. In most cases, however, delirium results from widespread disturbances in cortical and subcortical regions of the brain. Electroencephalogram (EEG) usually reveals symmetric slowing, a nonspecific finding that supports diffuse cerebral dysfunction.

Multiple neurotransmitter abnormalities, proinflammatory factors, and specific genes likely play a role in the pathogenesis of delirium. Deficiency of acetylcholine may play a key role, and medications with anticholinergic properties can commonly precipitate delirium. As noted earlier, patients with preexisting dementia are particularly susceptible to episodes of delirium. Alzheimer's disease (Chap. 431), dementia with Lewy bodies (Chap. 434), and Parkinson's disease dementia (Chap. 435) are all associated with cholinergic deficiency due to degeneration of acetylcholine-producing neurons in the basal forebrain. In addition, other neurotransmitters are also likely to be involved in this diffuse cerebral disorder. For example, increases in dopamine can lead to delirium, and patients with Parkinson's disease treated with dopaminergic medications can develop a delirium-like state that features visual hallucinations, fluctuations, and confusion.

Not all individuals exposed to the same insult will develop signs of delirium. A low dose of an anticholinergic medication may have no cognitive effects on a healthy young adult but produce a florid delirium in an elderly person with known underlying dementia, although even healthy young persons develop delirium with very high doses of anticholinergic medications. This concept of delirium developing as the result of an insult in predisposed individuals is currently the most widely accepted pathogenic construct. Therefore, if a previously healthy individual with no known history of cognitive illness develops delirium in the setting of a relatively minor insult such as elective surgery or hospitalization, an unrecognized underlying neurologic illness such as a neurodegenerative disease, multiple previous strokes, or another diffuse cerebral cause should be considered. In this context, delirium can be viewed as a "stress test for the brain" whereby exposure to known inciting factors such as systemic infection and offending drugs can unmask a decreased cerebral reserve and herald a serious underlying and potentially treatable illness. New blood-based biomarkers for specific dementias may soon be available to help predict people at risk for delirium before surgical procedures or hospitalization.

APPROACH TO THE PATIENT

Delirium

Because the diagnosis of delirium is clinical and is made at the bedside, a careful history and physical examination are necessary in evaluating patients with possible confusional states. Screening tools can aid physicians and nurses in identifying patients with delirium, including the Confusion Assessment Method (CAM); the Nursing Delirium Screening Scale (NuDESC); the Organic Brain Syndrome Scale; the Delirium Rating Scale; and, in the ICU, the ICU version of the CAM and the Delirium Detection Score. Using the well-validated CAM, a diagnosis of delirium is made if there is (1) an acute onset and fluctuating course and (2) inattention accompanied by either (3) disorganized thinking or (4) an altered level of consciousness (Table 27-1). These scales may not identify the full spectrum of patients with delirium, and all patients who are acutely confused should be presumed delirious regardless of their presentation due to the wide variety of possible clinical features. A course that fluctuates over hours or days and may worsen at night (termed *sundowning*) is typical but not essential for the diagnosis. Observation will usually reveal an altered level of consciousness or a deficit of attention. Other features that are sometimes present include alteration of sleep-wake cycles, thought disturbances such as hallucinations or delusions, autonomic instability, and changes in affect.

HISTORY

It may be difficult to elicit an accurate history in delirious patients who have altered levels of consciousness or impaired attention. Information from a collateral source such as a spouse or another family member is therefore invaluable. The three most important pieces of history are the patient's baseline cognitive function, the time course of the present illness, and current medications.

Premorbid cognitive function can be assessed through the collateral source or, if needed, via a review of outpatient records. Delirium by definition represents a change that is relatively acute and usually developing over hours to days, from a cognitive baseline. An acute confusional state is nearly impossible to diagnose without some knowledge of baseline cognitive function. Without

TABLE 27-1 The Confusion Assessment Method (CAM) Diagnostic Algorithm^a

The diagnosis of delirium requires the presence of features 1 and 2 and either feature 3 or 4.

Feature 1. Acute Onset and Fluctuating Course

This feature is satisfied by positive responses to the following questions: Is there evidence of an acute change in mental status from the patient's baseline? Did the (abnormal) behavior fluctuate during the day, that is, tend to come and go, or did it increase and decrease in severity?

Feature 2. Inattention

This feature is satisfied by a positive response to the following question: Did the patient have difficulty focusing attention, for example, being easily distractible, or have difficulty keeping track of what was being said?

Feature 3. Disorganized Thinking

This feature is satisfied by a positive response to the following question: Was the patient's thinking disorganized or incoherent, such as rambling or irrelevant conversation, unclear or illogical flow of ideas, or unpredictable switching from subject to subject?

Feature 4. Altered Level of Consciousness

This feature is satisfied by any answer other than "alert" to the following question: Overall, how would you rate the patient's level of consciousness: alert (normal), vigilant (hyperalert), lethargic (drowsy, easily aroused), stupor (difficult to arouse), or coma (unarousable)?

^aInformation is usually obtained from a reliable reporter, such as a family member, caregiver, or nurse.

Source: From Annals of Internal Medicine, SK Inouye et al: Clarifying confusion: The Confusion Assessment Method. A new method for detection of delirium. 113(12):941, 1990. Copyright © 1990 American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.

this information, many patients with dementia or longstanding depression may be mistaken as delirious during a single initial evaluation. Patients with a more hypoactive, apathetic presentation with psychomotor slowing may be identified as being different from baseline only through conversations with family members. A number of validated instruments have been shown to diagnose cognitive dysfunction accurately using a collateral source, including the modified Blessed Dementia Rating Scale and the Clinical Dementia Rating (CDR). Baseline cognitive impairment is common in patients with delirium. Even when no such history of cognitive impairment is elicited, there should still be a high suspicion for a previously unrecognized underlying neurologic disorder.

Establishing the time course of cognitive change is important not only to make a diagnosis of delirium but also to correlate the onset of the illness with potentially treatable etiologies such as recent medication changes or symptoms of systemic infection.

Medications remain a common cause of delirium, especially compounds with anticholinergic or sedative properties. It is estimated that nearly one-third of all cases of delirium are secondary to medications, especially in the elderly. Medication histories should include all prescription as well as over-the-counter and herbal substances taken by the patient and any recent changes in dosing or formulation, including substitution of generics for brand-name medications.

Other important elements of the history include screening for symptoms of organ failure or systemic infection, which often contributes to delirium in the elderly. A history of illicit drug use, alcoholism, or toxin exposure is common in younger delirious patients. Finally, asking the patient and collateral source about other symptoms that may accompany delirium, such as depression, may help identify potential therapeutic targets.

PHYSICAL EXAMINATION

The general physical examination in a delirious patient should include careful screening for signs of infection such as fever, tachypnea, pulmonary consolidation, heart murmur, and meningismus. The patient's fluid status should be assessed; both dehydration and fluid overload with resultant hypoxemia have been associated with delirium, and each is usually easily rectified. The appearance of the skin can be helpful, showing jaundice in hepatic encephalopathy, cyanosis in hypoxemia, or needle tracks in patients using intravenous drugs.

The neurologic examination requires a careful assessment of mental status. Patients with delirium often present with a fluctuating course; therefore, the diagnosis can be missed when one relies on a single time point of evaluation. For patients who worsen in the evening (sundowning), assessment only during morning rounds may be falsely reassuring.

An altered level of consciousness ranging from hyperarousal to lethargy to coma is present in most patients with delirium and can be assessed easily at the bedside. In a patient with a relatively normal level of consciousness, a screen for an attentional deficit is in order, because this deficit is the classic neuropsychological hallmark of delirium. Attention can be assessed while taking a history from the patient. Tangential speech, a fragmentary flow of ideas, or inability to follow complex commands often signifies an attentional problem. There are formal neuropsychological tests to assess attention, but a simple bedside test of digit span forward is quick and fairly sensitive. In this task, patients are asked to repeat successively longer random strings of digits beginning with two digits in a row, said to the patient at one per second intervals. Healthy adults can repeat a string of five to seven digits before faltering; a digit span of four or less usually indicates an attentional deficit unless hearing or language barriers are present, and many patients with delirium have digit spans of three or fewer digits.

More formal neuropsychological testing can be helpful in assessing a delirious patient, but it is usually too cumbersome and time-consuming in the inpatient setting. A Mini-Mental State

Examination (MMSE) provides information regarding orientation, language, and visuospatial skills ([Chap. 29](#)); however, performance of many tasks on the MMSE, including the spelling of "world" backward and serial subtraction of digits, will be impaired by delirious patients' attentional deficits, rendering the test unreliable.

The remainder of the screening neurologic examination should focus on identifying new focal neurologic deficits. Focal strokes or mass lesions in isolation are rarely the cause of delirium, but patients with underlying extensive cerebrovascular disease or neurodegenerative conditions may not be able to cognitively tolerate even relatively small new insults. Patients should be screened for other signs of neurodegenerative conditions such as parkinsonism, which is seen not only in idiopathic Parkinson's disease but also in other dementing conditions including Alzheimer's disease, dementia with Lewy bodies, and progressive supranuclear palsy. The presence of multifocal myoclonus or asterixis on the motor examination is nonspecific but usually indicates a metabolic or toxic etiology of the delirium.

ETIOLOGY

Some etiologies can be easily discerned through a careful history and physical examination, whereas others require confirmation with laboratory studies, imaging, or other ancillary tests. A large, diverse group of insults can lead to delirium, and the cause in many patients is multifactorial. Common etiologies are listed in [Table 27-2](#).

Prescribed, over-the-counter, and herbal medications all can precipitate delirium. Drugs with anticholinergic properties, narcotics, and benzodiazepines are particularly common offenders, but nearly any compound can lead to cognitive dysfunction in a predisposed patient. Whereas an elderly patient with baseline dementia may become delirious upon exposure to a relatively low dose of a medication, in less susceptible individuals, delirium occurs only with very high doses of the same medication. This observation emphasizes the importance of correlating the timing of recent medication changes, including dose and formulation, with the onset of cognitive dysfunction.

In younger patients, illicit drugs and toxins are common causes of delirium. In addition to more classic drugs of abuse, the availability of "bath salts," synthetic cannabis ([Chap. 455](#)), methylenedioxymethamphetamine (MDMA, ecstasy), γ -hydroxybutyrate (GHB), and the phencyclidine (PCP)-like agent ketamine has led to an increase in delirious young persons presenting to acute care settings ([Chap. 457](#)). Many common prescription drugs such as oral narcotics and benzodiazepines are often abused and readily available on the street. Alcohol abuse leading to high serum levels causes confusion, but more commonly, it is withdrawal from alcohol that leads to a hyperactive delirium ([Chap. 453](#)). Alcohol and benzodiazepine withdrawal should be considered in all cases of delirium, including in the elderly, because even patients who drink only a few servings of alcohol every day can experience relatively severe withdrawal symptoms upon hospitalization.

Metabolic abnormalities such as electrolyte disturbances of sodium, calcium, magnesium, or glucose can cause delirium, and mild derangements can lead to substantial cognitive disturbances in susceptible individuals. Other common metabolic etiologies include liver and renal failure, hypercarbia and hypoxemia, vitamin deficiencies of thiamine and B₁₂, autoimmune disorders including central nervous system (CNS) vasculitis, and endocrinopathies such as thyroid and adrenal disorders.

Systemic infections often cause delirium, especially in the elderly. A common scenario involves the development of an acute cognitive decline in the setting of a urinary tract infection in a patient with baseline dementia. Pneumonia, skin infections such as cellulitis, and frank sepsis also lead to delirium. This so-called septic encephalopathy, often seen in the ICU, is probably due to the release of proinflammatory cytokines and their diffuse cerebral effects. CNS infections such as meningitis, encephalitis, and abscess are less common etiologies of delirium, as are cases of autoimmune or

TABLE 27-2 Differential Diagnosis of Delirium**Toxins**

Prescription medications: especially those with anticholinergic properties, narcotics, and benzodiazepines
 Drugs of abuse: alcohol intoxication and alcohol withdrawal, opiates, ecstasy, LSD, GHB, PCP, ketamine, cocaine, "bath salts," marijuana and its synthetic forms
 Poisons: inhalants, carbon monoxide, ethylene glycol, pesticides

Metabolic Conditions

Electrolyte disturbances: hypoglycemia, hyperglycemia, hyponatremia, hypernatremia, hypercalcemia, hypocalcemia, hypomagnesemia
 Hypothermia and hyperthermia
 Pulmonary failure: hypoxemia and hypercarbia
 Liver failure/hepatic encephalopathy
 Renal failure/uremia
 Cardiac failure
 Vitamin deficiencies: B₁₂, thiamine, folate, niacin
 Dehydration and malnutrition
 Anemia

Infections

Systemic infections: urinary tract infections, pneumonia, skin and soft tissue infections, sepsis
 CNS infections: meningitis, encephalitis, brain abscess

Endocrine Conditions

Hyperthyroidism, hypothyroidism
 Hyperparathyroidism
 Adrenal insufficiency

Cerebrovascular Disorders

Global hypoperfusion states
 Hypertensive encephalopathy
 Focal ischemic strokes and hemorrhages (rare): especially nondominant parietal and thalamic lesions

Autoimmune Disorders

CNS vasculitis
 Cerebral lupus
 Neurologic paraneoplastic and autoimmune encephalitis

Seizure-Related Disorders

Nonconvulsive status epilepticus
 Intermittent seizures with prolonged postictal states

Neoplastic Disorders

Diffuse metastases to the brain
 Gliomatosis cerebri
 Carcinomatous meningitis
 CNS lymphoma

Hospitalization

Terminal end-of-life delirium

Abbreviations: CNS, central nervous system; GHB, γ -hydroxybutyrate; LSD, lysergic acid diethylamide; PCP, phencyclidine.

paraneoplastic encephalitis; however, in light of the high morbidity and mortality rates associated with these conditions when they are not treated, clinicians must always maintain a high index of suspicion.

In some susceptible individuals, exposure to the unfamiliar environment of a hospital itself can contribute to delirium. This etiology usually occurs as part of a multifactorial delirium and should be considered a diagnosis of exclusion after all other causes have been thoroughly investigated. Many primary prevention and treatment strategies for delirium involve relatively simple methods to address the aspects of the inpatient setting that are most confusing.

Cerebrovascular etiologies of delirium are usually due to global hypoperfusion in the setting of systemic hypotension from heart failure, septic shock, dehydration, or anemia. Focal strokes in the right parietal lobe and medial dorsal thalamus rarely can lead to a delirious state. A more common scenario involves a new focal stroke or hemorrhage causing confusion in a patient who has decreased cerebral reserve. In these individuals, it is sometimes difficult to distinguish between cognitive dysfunction resulting from the new neurovascular insult itself and delirium due to the infectious, metabolic, and pharmacologic complications that can accompany hospitalization after stroke.

Because a fluctuating course often is seen in delirium, intermittent seizures may be overlooked when one is considering potential etiologies. Both nonconvulsive status epilepticus and recurrent focal or generalized seizures followed by postictal confusion can cause delirium; EEG remains essential for this diagnosis and should be considered whenever the etiology of delirium remains unclear following initial workup. Seizure activity spreading from an electrical focus in a mass or infarct can explain global cognitive dysfunction caused by relatively small lesions.

It is extremely common for patients to experience delirium at the end of life in palliative care settings. This condition must be identified and treated aggressively because it is an important cause of patient and family discomfort at the end of life. It should be remembered that these patients also may be suffering from more common etiologies of delirium such as systemic infection.

LABORATORY AND DIAGNOSTIC EVALUATION

A cost-effective approach allows the history and physical examination to guide further tests. No single algorithm will fit all delirious patients due to the staggering number of potential etiologies, but one stepwise approach is detailed in **Table 27-3**. If a clear precipitant such as an offending medication is identified, further testing may not be required. If, however, no likely etiology is uncovered with initial evaluation, an aggressive search for an underlying cause should be initiated.

Basic screening labs, including a complete blood count, electrolyte panel, and tests of liver and renal function, should be obtained in all patients with delirium. In elderly patients, screening for systemic infection, including chest radiography, urinalysis and culture, and possibly blood cultures, is important. In younger individuals, serum and urine drug and toxicology screening may be appropriate earlier in the workup. Additional laboratory tests addressing other autoimmune, endocrinologic, metabolic, and infectious etiologies should be reserved for patients in whom the diagnosis remains unclear after initial testing.

Multiple studies have demonstrated that brain imaging in patients with delirium is often unhelpful. If, however, the initial workup is unrevealing, most clinicians quickly move toward imaging of the brain to exclude structural causes. A noncontrast computed tomography (CT) scan can identify large masses and hemorrhages but is otherwise unlikely to help determine an etiology of delirium. The ability of magnetic resonance imaging (MRI) to identify most acute ischemic strokes as well as to provide neuroanatomic detail that gives clues to possible infectious, inflammatory, neurodegenerative, and neoplastic conditions makes it the test of choice. Because MRI techniques are limited by availability, speed of imaging, patient's cooperation, and contraindications, many clinicians begin with CT scanning and proceed to MRI if the etiology of delirium remains elusive.

Lumbar puncture (LP) must be obtained immediately after neuroimaging for all patients in whom CNS infection is suspected. Spinal fluid examination can also be useful in identifying autoimmune, other inflammatory, and neoplastic conditions. As a result, LP should be considered in any delirious patient with a negative workup. EEG remains invaluable if seizures are considered or if there is no cause readily identified.

TABLE 27-3 Stepwise Evaluation of a Patient with Delirium**Initial Evaluation**

History with special attention to medications (including over-the-counter and herbals)
General physical examination and neurologic examination
Complete blood count
Electrolyte panel including calcium, magnesium, phosphorus
Liver function tests, including albumin
Renal function tests

First-Tier Further Evaluation Guided by Initial Evaluation

Systemic infection screen
Urinalysis and culture
Chest radiograph
Blood cultures
Electrocardiogram
Arterial blood gas
Serum and/or urine toxicology screen (perform earlier in young persons)
Brain imaging with MRI with diffusion and gadolinium (preferred) or CT
Suspected CNS infection or other inflammatory disorder: lumbar puncture after brain imaging
Suspected seizure-related etiology: electroencephalogram (EEG) (if high suspicion, should be performed immediately)

Second-Tier Further Evaluation

Vitamin levels: B₁₂, folate, thiamine
Endocrinologic laboratories: thyroid-stimulating hormone (TSH) and free T₄; cortisol
Serum ammonia
Sedimentation rate
Autoimmune serologies: antinuclear antibodies (ANA), complement levels; p-ANCA, c-ANCA, consider paraneoplastic/autoimmune encephalitis serologies
Infectious serologies: rapid plasmin reagent (RPR); fungal and viral serologies if high suspicion; HIV antibody
Lumbar puncture (if not already performed)
Brain MRI with and without gadolinium (if not already performed)

Abbreviations: c-ANCA, cytoplasmic antineutrophil cytoplasmic antibody; CNS, central nervous system; CT, computed tomography; MRI, magnetic resonance imaging; p-ANCA, perinuclear antineutrophil cytoplasmic antibody.

TREATMENT**Delirium**

Management of delirium begins with treatment of the underlying inciting factor (e.g., patients with systemic infections should be given appropriate antibiotics, and underlying electrolyte disturbances should be judiciously corrected). These treatments often lead to prompt resolution of delirium. Blindly targeting the symptoms of delirium pharmacologically only serves to prolong the time patients remain in the confused state and may mask important diagnostic information.

Relatively simple methods of supportive care can be highly effective (**Fig. 27-1**). Reorientation by the nursing staff and family combined with visible clocks, calendars, and outside-facing windows can reduce confusion. Sensory isolation should be prevented by providing glasses and hearing aids to patients who need them. Sundowning can be addressed to a large extent through vigilance to appropriate sleep-wake cycles. During the day, a well-lit room should be accompanied by activities or exercises to prevent napping. At night, a quiet, dark environment with limited interruptions by staff can assure proper rest; melatonin can be considered before bed to promote sleep. These sleep-wake cycle interventions are especially important in the ICU setting as the usual constant 24-h activity commonly provokes delirium. Attempting to mimic the home environment as much as possible also has been shown to help treat and even prevent delirium. Visits from friends and

PROMOTE AM • WAKEFULNESS

Shades up. Lights on.	Write date and staff names on board to orient patient.	Patient out of bed to chair for all 3 meals. Ask for assistance if you need help.	Walk patient 3x/day. Engage patient in conversation.

Each visit, introduce yourself; remind patient where they are, what day and time it is.	Patient is wearing hearing aids/glasses (if needed) to hear and see appropriately.	Provide activities like games and reading materials to keep patient's mind active while awake.	Make sure your patient has water within reach at all times. Dehydration is the #1 complaint in the hospital!

Make sure family members have been provided the pamphlet about delirium and discuss any questions they have. It is ok to refer to the nurse or doctor if you are unsure.	Discuss with the nurse at each shift if the patient truly needs the following: nasal cannula on their nose, Foley catheter, telemetry, and CPO. These "tethers" make it difficult for the patient to move and can contribute to confusion.

A**PROMOTE PM • SLEEP**

Shades closed. Lights off. TV off. Make room as dark and quiet as possible.	Minimize caffeine intake.	Offer eye mask, ear plugs to help with sleep.

Group your nighttime tasks so that you are entering the room and waking the patient as few times as possible.	If you communicate with the patient during the night, make sure glasses and hearing aids are on. Remember to introduce yourself, remind the patient where they are.

B

FIGURE 27-1 Delirium management and prevention: a checklist for hospitalized patients. Effective management of delirium relies on broad efforts to promote wakefulness (A) and sleep (B). CPO, continuous pulse oximetry.

family throughout the day minimize the anxiety associated with the constant flow of new faces of staff and physicians. Allowing hospitalized patients to have access to home bedding, clothing, and nightstand objects makes the hospital environment less foreign and therefore less confusing. Simple standard nursing practices such as maintaining proper nutrition and volume status as well as managing pain, incontinence, and skin breakdown also help alleviate discomfort and resulting confusion.

In some instances, patients pose a threat to their own safety or to the safety of staff members, and acute management is required. Bed alarms and personal sitters are more effective and much less disorienting than physical restraints. Chemical restraints should be avoided, but when necessary, very-low-dose typical or atypical antipsychotic medications administered on an as-needed basis can be used, recognizing that clinical trials have consistently shown that these medications are ineffective in treating delirium. Therefore, they should be reserved for patients who display severe agitation and significant potential to harm themselves or staff. The association of antipsychotic use in the elderly with increased mortality rates underscores the importance of using these medications judiciously and only as a last resort. Benzodiazepines often worsen

confusion through their sedative properties. Although many clinicians use benzodiazepines to treat acute confusion, their use should be limited to cases in which delirium is caused by alcohol or benzodiazepine withdrawal.

■ PREVENTION

In light of the high morbidity associated with delirium and the tremendously increased health care costs that accompany it, development of an effective strategy to prevent delirium in hospitalized patients is extremely important. Successful identification of high-risk patients is the first step, followed by initiation of appropriate interventions. Increasingly, hospitals are using nursing or physician-administered tools to screen for high-risk individuals, triggering simple standardized protocols used to manage risk factors for delirium, including sleep-wake cycle reversal, immobility, visual impairment, hearing impairment, sleep deprivation, and dehydration. No specific medications have been definitively shown to be effective for delirium prevention, including trials of cholinesterase inhibitors and antipsychotic agents. Melatonin and its agonist ramelteon have shown some promising results in small preliminary trials. Recent studies in the ICU have focused both on identifying sedatives, such as dexmedetomidine, that are less likely to lead to delirium in critically ill patients and on developing protocols for daily awakenings in which infusions of sedative medications are interrupted and the patient is reoriented by the staff. All hospitals and health care systems should work toward decreasing the incidence of delirium and promptly recognizing and treating the disorder when it occurs.

■ FURTHER READING

- BROWN EG et al: Evaluation of a multicomponent pathway to address inpatient delirium on a neurosciences ward. *BMC Health Serv Res* 18:106, 2018.
- CONSTANTIN JM et al: Efficacy and safety of sedation with dexmedetomidine in critical care patients: A meta-analysis of randomized controlled trials. *Anaesth Crit Care Pain Med* 35:7, 2016.
- GIRARD TD et al: Haloperidol and ziprasidone for treatment of delirium in critical illness. *N Engl J Med* 379:2506, 2018.
- GOLDBERG TE et al: Association of delirium with long-term cognitive decline: A meta-analysis. *JAMA Neurol* 77:1, 2020.
- HATTA K et al: Preventive effects of ramelteon on delirium: A randomized placebo-controlled trial. *JAMA Psychiatry* 71:397, 2014.

preferable to use of ambiguous terms such as lethargy, semicoma, or obtundation.

Several conditions that render patients unresponsive and simulate coma are considered separately because of their special significance. The *vegetative state* signifies an awake-appearing but nonresponsive state, usually encountered in a patient who has emerged from coma. In the vegetative state, the eyelids may open periodically, giving the appearance of wakefulness. Respiratory and autonomic functions are retained. Yawning, coughing, swallowing, and limb and head movements persist, but there are few, if any, meaningful responses to the external and internal environment. There are typically accompanying signs that indicate extensive damage in both cerebral hemispheres, e.g., decerebrate or decorticate limb posturing and absent responses to visual stimuli (see below). In the closely related but less severe *minimally conscious state*, the patient displays rudimentary vocal or motor behaviors, often spontaneous, but sometimes in response to touch, visual stimuli, or command. Cardiac arrest with cerebral hypoperfusion and head trauma are the most common causes of the vegetative and minimally conscious states (**Chap. 307**).

The prognosis for regaining meaningful mental faculties once the vegetative state has supervened for several months is poor, and after a year, almost nil; hence the term *persistent vegetative state*. Most reports of dramatic recovery, when investigated carefully, are found to yield to the usual rules for prognosis, but there have been rare instances in which recovery has occurred to a severely disabled condition and, in rare childhood cases, to an even better state. Patients in the minimally conscious state carry a better prognosis for some recovery compared to those in a persistent vegetative state, but even in these patients, dramatic recovery after 12 months is unusual.

The possibility of incorrectly attributing meaningful behavior to patients in the vegetative and minimally conscious states creates problems and anguish for families and physicians. The question of whether some of these patients have the capability for cognition has been investigated by functional MRI and electroencephalogram (EEG) studies that have demonstrated cerebral activation that is temporally consistent in response to verbal and other stimuli, as discussed in more detail below. This finding suggests at a minimum that some of these patients could in the future be able to communicate their needs using technological advances and that further research could shed light on treatment approaches targeting areas of the brain and their connections that seem to be preserved in individual patients.

Several syndromes that affect alertness are prone to be misinterpreted as stupor or coma, and clinicians should be aware of these pitfalls when diagnosing coma at the bedside. Akinetic mutism refers to a partially or fully awake state in which the patient remains virtually immobile and mute but can form impressions and think, as demonstrated by later recounting of events. This condition results from damage in the regions of the medial thalamic nuclei or the frontal lobes (particularly lesions situated deeply or on the orbitofrontal surfaces) or from extreme hydrocephalus. The term *abulia* describes a milder form of akinetic mutism characterized by mental and physical slowness and diminished ability to initiate activity. It is also usually the result of damage to the medial frontal lobes and their connections (**Chap. 30**).

Catatonia is a hypomobile and mute syndrome that occurs usually as part of a major psychosis, typically schizophrenia or major depression. Catatonic patients make few voluntary or responsive movements, although they blink, swallow, and may not appear distressed. There are nevertheless signs that the patient is responsive, although it takes a careful examination to demonstrate these features. For example, eyelid elevation is actively resisted, blinking occurs in response to a visual threat, and the eyes move concomitantly with head rotation, all of which are inconsistent with the presence of a brain lesion causing unresponsiveness. The limbs may retain postures in which they have been placed by the examiner ("waxy flexibility," or catalepsy). With recovery from catatonia, patients often have some memory of events that occurred during their stupor. Catatonia is superficially similar to akinetic mutism, but clinical evidence of cerebral damage such as hyperreflexia and hypertonicity of the limbs is lacking in the former. The special problem of coma in brain death is discussed below.

28

Coma

S. Andrew Josephson, Allan H. Ropper,
Stephen L. Hauser

Coma is among the most common neurologic emergencies encountered general medicine and requires an organized approach. It accounts for a substantial portion of admissions to emergency wards and occurs on all hospital services.

There exists a continuum of states of reduced alertness, the most severe form being coma, defined as a deep sleeplike state with eyes closed, from which the patient cannot be aroused. Stupor refers to a lower threshold for arousability, in which the patient can be transiently awakened by vigorous stimuli, accompanied by motor behavior that leads to avoidance or withdrawal from noxious stimuli. Drowsiness simulates light sleep and is characterized by easy arousal that may persist for brief periods. Stupor and drowsiness are usually accompanied by some degree of confusion when the patient is alerted (**Chap. 27**). A precise narrative description of the level of arousal and of the type of responses evoked by various stimuli as observed at the bedside is

The locked-in state describes a type of pseudocoma in which an awake but paralyzed patient has no means of producing speech or voluntary limb movement but retains voluntary vertical eye movements and lid elevation, thus allowing the patient to communicate. The pupils are normally reactive. The usual cause is an infarction (e.g., basilar artery thrombosis) or hemorrhage of the bilateral ventral pons that transects all descending motor (corticospinal and corticobulbar) pathways. Another awake but de-efferent state occurs as a result of total paralysis of the musculature in severe cases of neuromuscular weakness such as in Guillain-Barré syndrome (**Chap. 447**), critical illness neuropathy (**Chap. 307**), or pharmacologic neuromuscular blockade.

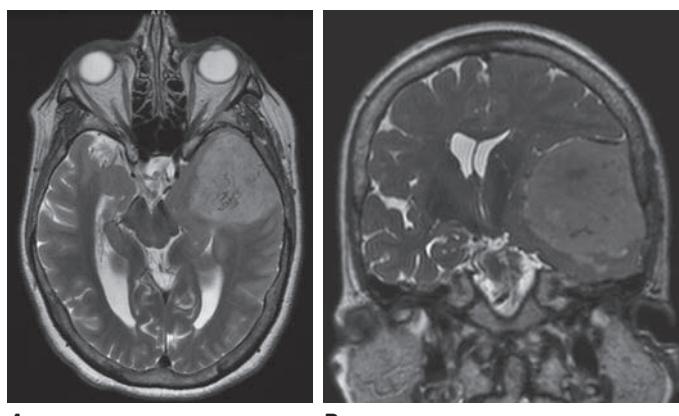
THE ANATOMY AND PHYSIOLOGY OF COMA

Almost all instances of coma can be traced to either (1) widespread abnormalities of the cerebral hemispheres or (2) reduced activity of the thalamocortical alerting system, the reticular activating system (RAS), which is an assemblage of neurons located diffusely in the upper brainstem and thalamus. The proper functioning of this system, its ascending projections to the cortex, and the cortex itself are required to maintain alertness and coherence of thought. In addition to structural damage to either or both of these systems, suppression of reticulocerebral function commonly occurs by drugs, toxins, or metabolic derangements such as hypoglycemia, anoxia, uremia, and hepatic failure, or by seizures; these types of metabolic causes of coma are far more common than structural injuries.

Coma Due to Cerebral Mass Lesions and Herniation Syndromes

Syndromes The skull prevents outward expansion of the brain, and infoldings of the dura create compartments that restrict displacement of brain tissue within the cranium. The two cerebral hemispheres are separated by the falx and the anterior and posterior fossae by the tentorium. Herniation refers to displacement of brain tissue by an intracerebral or overlying mass into a contiguous compartment that it normally does not occupy. Coma from mass lesions, and many of its associated signs, are attributable to these tissue shifts, and certain clinical features are characteristic of specific configurations of herniation (**Fig. 28-1**).

In the most common form of herniation, brain tissue is displaced from the supratentorial to the infratentorial compartment through the tentorial opening, referred to as transtentorial herniation. The cause is often a mass hemispherical lesion, with accompanying contralateral hemiparesis. Uncal transtentorial herniation refers to impaction of the anterior medial temporal gyrus (the uncus) into the tentorial opening just anterior to and adjacent to the midbrain (**Fig. 28-1A**). The uncus can compress the third nerve as the nerve traverses the subarachnoid space, causing enlargement of the ipsilateral pupil as the first sign (the fibers subserving parasympathetic pupillary function are located



A B

FIGURE 28-2 Axial (A) and coronal (B) T2-weighted magnetic resonance images from a stuporous patient with a left third nerve palsy from a large left-sided meningioma. **A.** The upper midbrain is compressed and displaced horizontally away from the mass, and there is transtentorial herniation of the medial temporal lobe structures, including the uncus. **B.** The lateral ventricle opposite to the mass has become enlarged as a result of compression of the third ventricle.

peripherally in the nerve). The coma that typically follows is due to lateral displacement of the midbrain (and therefore the RAS) against the opposite tentorial edge by the displaced parahippocampal gyrus (**Fig. 28-2**), compressing the opposite cerebral peduncle and producing a Babinski sign and ipsilateral hemiparesis (the Kernohan-Woltman sign). Herniation may also compress the anterior and posterior cerebral arteries as they pass over the tentorial reflections, with resultant brain infarction. These distortions may also entrap portions of the ventricular system, causing hydrocephalus.

Central transtentorial herniation denotes a symmetric downward movement of the thalamic structures through the tentorial opening with compression of the upper midbrain (**Fig. 28-1B**). Miotic pupils and drowsiness are the heralding signs, in contrast to a unilaterally enlarged pupil of the uncal syndrome. Both uncal and central transtentorial herniations cause progressive compression of the brainstem and RAS, with initial damage to the midbrain, then the pons, and finally the medulla. The result is an approximate sequence of neurologic signs that corresponds to each affected level, with respiratory centers in the brainstem often spared until late in the herniation syndrome. Other forms of herniation include transfalcial herniation (displacement of the cingulate gyrus under the falx and across the midline, **Fig. 28-1C**) and foraminal herniation (downward forcing of the cerebellar tonsils into the foramen magnum, **Fig. 28-1D**), which causes early compression of the medulla, respiratory arrest, and death.

Coma Due to Metabolic, Drug, and Toxic Disorders Many systemic metabolic abnormalities cause coma by interrupting the delivery of energy substrates (e.g., oxygen, glucose) or by altering neuronal excitability (drugs and alcohol, anesthesia, and epilepsy). These are the most common causes of coma in large case series. The metabolic abnormalities that produce coma may, in milder forms, induce a confusional state (metabolic encephalopathy) in which clouded consciousness and coma are in a continuum.

Cerebral neurons are dependent on cerebral blood flow (CBF) and the delivery of oxygen and glucose. Brain stores of glucose are able to provide energy for ~2 min after blood flow is interrupted, and oxygen stores last 8–10 s after the cessation of blood flow. Simultaneous hypoxia and ischemia exhaust glucose more rapidly. The EEG rhythm in these circumstances becomes diffusely slowed, typical of metabolic encephalopathies, and as substrate delivery worsens, eventually brain electrical activity ceases.

Unlike hypoxia-ischemia, which first causes a metabolic encephalopathy due to reduced energy substrate but ultimately causes neuronal destruction, most metabolic disorders such as hypoglycemia, hyponatremia, hyperosmolarity, hypercapnia, hypercalcemia, and hepatic and renal failure cause no or only minor neuropathologic changes in the

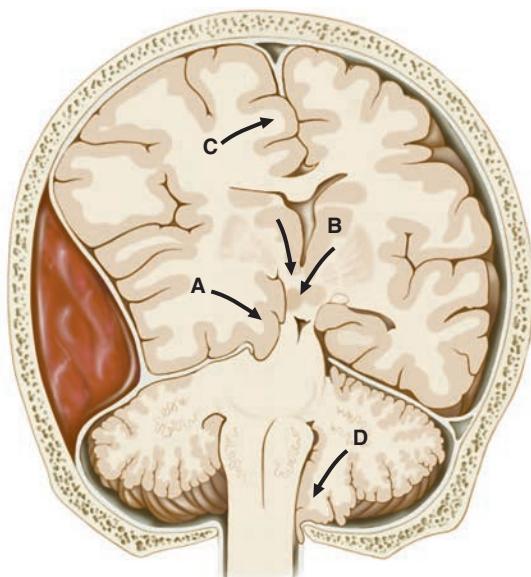


FIGURE 28-1 Types of cerebral herniation: (A) uncal; (B) central; (C) transfalcial; and (D) foraminal.

brain. The reversible effects of these conditions are not fully understood but may result from impaired energy supplies, changes in ion fluxes across neuronal membranes, and neurotransmitter abnormalities. In hepatic encephalopathy (HE), high ammonia concentrations lead to increased synthesis of glutamine in astrocytes and osmotic swelling of the cells, mitochondrial energy failure, production of reactive nitrogen and oxygen species, increases in the inhibitory neurotransmitter GABA, and synthesis of putative “false” neurotransmitters. Over time, development of a diffuse astrocytosis is typical of chronic HE. Which, if any, of these is responsible for coma is not known.

The mechanism of the encephalopathy of renal failure is also uncertain and likely to be multifactorial; unlike ammonia, urea does not produce central nervous system (CNS) depression. Contributors to uremic encephalopathy may include accumulation of neurotoxic substances such as creatinine, guanidine, and related compounds; depletion of catecholamines; altered glutamate and GABA tone; increases in brain calcium; inflammation with disruption of the blood-brain barrier; and frequent coexisting vascular disease.

Coma and seizures are common accompaniments of large shifts in sodium and water balance in the brain. These changes in osmolarity arise from systemic medical disorders, including diabetic ketoacidosis, the nonketotic hyperosmolar state, and hyponatremia from any cause (e.g., water intoxication, excessive secretion of antidiuretic hormone, or atrial natriuretic peptides). Sodium levels <125 mmol/L, especially if achieved quickly, induce confusion, and levels <119 mmol/L are typically associated with coma and convulsions. In hyperosmolar coma, the serum osmolarity is generally >350 mosmol/L. Hypercapnia depresses the level of consciousness in proportion to the rise in carbon dioxide (CO_2) in the blood. In all of these metabolic encephalopathies, the degree of neurologic change depends on the rapidity with which the serum changes occur. The pathophysiology of other metabolic encephalopathies such as those due to hypercalcemia, hypothyroidism, vitamin B₁₂ deficiency, and hypothermia are incompletely understood but must reflect derangements of CNS biochemistry, membrane function, or neurotransmitters.

Comas due to drugs and toxins are typically reversible and leave no residual damage provided there has not been hypoxia or severe hypotension. Many drugs and toxins are capable of depressing nervous system function. Some produce coma by affecting both the RAS and the cerebral cortex. The combination of cortical and brainstem signs, which occurs occasionally in certain drug overdoses, may lead to an incorrect diagnosis of structural brainstem disease. Overdose of medications that have atropinic actions produces signs such as dilated pupils, tachycardia, and dry skin; opiate overdose produces pinpoint pupils <1 mm in diameter. Some drug intoxications, typified by barbiturates, can mimic all of the signs of brain death; thus, toxic etiologies should be excluded prior to making a diagnosis of brain death.

Epileptic Coma Generalized electrical seizures are associated with coma, even in the absence of motor convulsions (nonconvulsive status epilepticus). As a result, EEG monitoring is often used in the evaluation of unexplained coma to exclude this treatable etiology. The self-limited coma that follows a seizure, the postictal state, may be due to exhaustion of energy reserves or effects of locally toxic molecules that are the by-product of seizures. The postictal state produces continuous, generalized slowing of the background EEG activity similar to that of metabolic encephalopathies. It typically lasts for a few minutes but in some cases can be prolonged for hours or even rarely for days.

Coma Due to Widespread Structural Damage to the Cerebral Hemispheres This category, comprising several unrelated disorders, results from extensive bilateral structural cerebral damage. The clinical appearance simulates a metabolic encephalopathy. Hypoxia-ischemia is perhaps the best characterized form of this type of injury, in which it is not possible initially to distinguish the acute reversible effects of oxygen deprivation of the brain from the subsequent effects of anoxic neuronal damage. Similar cerebral damage may be produced by disorders that occlude widespread small blood vessels throughout the brain; examples include thrombotic thrombocytopenic purpura,

hyperviscosity, and cerebral malaria. Diffuse white matter damage from cranial trauma or inflammatory demyelinating diseases can cause a similar coma syndrome.

APPROACH TO THE PATIENT

Coma

A video examination of the comatose patient is shown in Chap. V4.

Acute respiratory and cardiovascular problems should be attended to prior to neurologic assessment. In most instances, a complete medical evaluation, except for vital signs, funduscopic, and examination for nuchal rigidity, may be deferred until the neurologic evaluation has established the severity and nature of coma. **The approach to the patient with coma from cranial trauma is discussed in Chap. 443.**

HISTORY

The cause of coma may be immediately evident as in cases of trauma, cardiac arrest, or observed drug ingestion. In the remainder, certain points are useful: (1) the circumstances and rapidity with which neurologic symptoms developed; (2) antecedent symptoms (confusion, weakness, headache, fever, seizures, dizziness, double vision, or vomiting); (3) the use of medications, drugs, or alcohol; and (4) chronic liver, kidney, lung, heart, or other medical disease. Direct interrogation of family, observers, and emergency medical technicians on the scene, in person or by telephone, is an important part of the evaluation when possible.

GENERAL PHYSICAL EXAMINATION

Signs of head trauma raise the possibility of coexisting spinal cord injury, and in such cases, immobilization of the cervical spine is essential to prevent further injury. Fever suggests a systemic infection, bacterial meningitis, encephalitis, heat stroke, neuroleptic malignant syndrome, malignant hyperthermia due to anesthetics, or anticholinergic drug intoxication. Only rarely is fever attributable to a lesion that has disturbed hypothalamic temperature-regulating centers (“central fever”), and this diagnosis should only be considered after an exhaustive search for other causes fails to reveal an explanation for fever. A slight elevation in temperature may follow vigorous convulsions. Hypothermia is observed with alcohol, barbiturate, sedative, or phenothiazine intoxication; hypoglycemia; peripheral circulatory failure; or extreme hypothyroidism. Hypothermia itself causes coma when the temperature is <31°C (87.8°F) regardless of the underlying etiology; less dramatically low body temperatures can also cause coma in some instances. Tachypnea may indicate systemic acidosis or pneumonia. Aberrant respiratory patterns that reflect brainstem disorders are discussed below. Marked hypertension suggests hypertensive encephalopathy, cerebral hemorrhage, large cerebral infarction, or head injury. Hypotension is characteristic of coma from alcohol or barbiturate intoxication, internal hemorrhage or myocardial infarction causing poor delivery of blood to the brain, sepsis, profound hypothyroidism, or Addisonian crisis. The fundoscopic examination can detect increased intracranial pressure (ICP) (papilledema), subarachnoid hemorrhage (subhyaloid hemorrhages), and hypertensive encephalopathy (exudates, hemorrhages, vessel-crossing changes, papilledema). Cutaneous petechiae suggest thrombotic thrombocytopenic purpura, meningococcemia, or a bleeding diathesis associated with an intracerebral hemorrhage. Cyanosis and reddish or anemic skin coloration are other indications of an underlying systemic disease or carbon monoxide as responsible for the coma.

NEUROLOGIC EXAMINATION

The patient should first be observed without intervention by the examiner. Spontaneously moving about the bed, reaching up toward the face, crossing legs, yawning, swallowing, coughing, and moaning reflect a drowsy state that is close to normal awkeness. Lack of restless movements on one side or an outturned leg suggests hemiplegia. Subtle, intermittent twitching movements of a foot, finger, or

facial muscle may be the only sign of seizures. Multifocal myoclonus usually indicates a metabolic disorder, particularly uremia, anoxia, drug intoxication, or rarely a prion disease ([Chap. 438](#)). In a drowsy and confused patient, bilateral asterixis is a sign of metabolic encephalopathy or drug intoxication.

Decorticate rigidity and decerebrate rigidity, or “posturing,” describe stereotyped arm and leg movements occurring spontaneously or elicited by sensory stimulation. Flexion of the elbows and wrists and supination of the arm (decorticate posturing) classically suggest bilateral damage rostral to the midbrain, whereas extension of the elbows and wrists with pronation (decerebrate posturing) indicates damage to motor tracts caudal to the midbrain. However, these localizations have been adapted from animal work and cannot be applied with precision to coma in humans. In fact, acute and widespread disorders of any type, regardless of location, frequently cause limb extension.

LEVEL OF AROUSAL

A sequence of increasingly intense stimuli is first used to determine the threshold for arousal and the motor response of each side of the body. The results of testing may vary from minute to minute, and serial examinations are useful. Tickling the nostrils with a cotton wisp is a moderate stimulus to arousal—all but deeply stuporous and comatose patients will move the head away and arouse to some degree. An even greater degree of responsiveness is present if the patient uses his hand to remove an offending stimulus. Pressure on bony prominences and pinprick stimulation, when necessary, are humane forms of noxious stimuli; pinching the skin causes ecchymoses and is generally not performed but may be useful in eliciting abduction withdrawal movements of the limbs. Posturing in response to noxious stimuli indicates severe damage to the corticospinal system, whereas abduction-avoidance movement of a limb is usually purposeful and denotes an intact corticospinal system. Posturing may also be unilateral and coexist with purposeful limb movements, reflecting incomplete damage to the motor system.

BRAINSTEM REFLEXES

Assessment of brainstem function is essential to localization of the lesion in coma ([Fig. 28-3](#)). Patients with preserved brainstem reflexes typically have a bihemispheric localization to coma, including toxic or drug intoxication, whereas patients with abnormal brainstem reflexes either have a lesion in the brainstem or a herniation syndrome from a cerebral mass lesion impacting the brainstem secondarily. The most important brainstem reflexes are pupillary size and reaction to light, spontaneous and elicited eye movements, corneal responses, and the respiratory pattern.

Pupillary Signs Pupillary reactions are examined with a bright, diffuse light. Reactive and round pupils of midsize (2.5–5 mm) essentially exclude upper midbrain damage, either primary or secondary to compression from herniation. A response to light may be difficult to appreciate in pupils <2 mm in diameter, and bright room lighting may mute pupillary reactivity. One enlarged (>6 mm) and poorly reactive pupil signifies compression of the third nerve from the effects of a cerebral mass above. Enlargement of the pupil contralateral to a hemispherical mass may occur but is infrequent. An oval and slightly eccentric pupil is a transitional sign that accompanies early midbrain-third nerve compression. The most extreme pupillary sign, bilaterally dilated and unreactive pupils, indicates severe midbrain damage, usually from compression by a supratentorial mass. Ingestion of drugs with anticholinergic activity, the use of mydriatic eye drops, nebulizer treatments, and direct ocular trauma are other causes of pupillary enlargement.

Reactive and bilaterally small (1–2.5 mm) but not pinpoint pupils are seen in metabolic encephalopathies or in deep bilateral hemispherical lesions such as hydrocephalus or thalamic hemorrhage. Even smaller reactive pupils (<1 mm) characterize opioid overdoses but also occur with extensive pontine hemorrhage. The response to naloxone and the presence of reflex eye movements (see

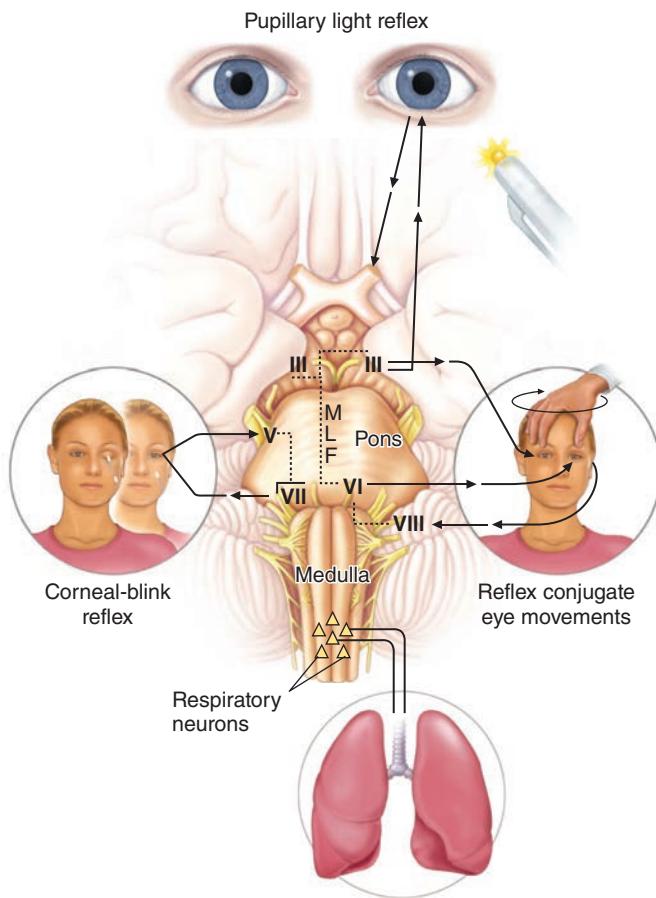


FIGURE 28-3 Examination of brainstem reflexes in coma. Midbrain and third nerve function are tested by pupillary reaction to light, pontine function by spontaneous and reflex eye movements and corneal responses, and medullary function by respiratory and pharyngeal responses. Reflex conjugate, horizontal eye movements are dependent on the medial longitudinal fasciculus (MLF) interconnecting the sixth and contralateral third nerve nuclei. Head rotation (oculocephalic reflex) or caloric stimulation of the labyrinths (oculovestibular reflex) elicits contraversive eye movements (for details, see text).

below) assist in distinguishing between these. Unilateral miosis in coma has been attributed to dysfunction of sympathetic efferents originating in the posterior hypothalamus and descending in the tegmentum of the brainstem to the cervical cord. It is an occasional finding in patients with a large cerebral hemorrhage that affects the thalamus.

Ocular Movements The eyes are first observed by elevating the lids and observing the resting position and spontaneous movements of the globes. Horizontal divergence of the eyes at rest is normal in drowsiness. As coma deepens, the ocular axes may become parallel again.

Spontaneous eye movements in coma often take the form of conjugate horizontal roving. This finding alone exonerates extensive damage in the midbrain and pons and has the same significance as normal reflex eye movements (see below). Conjugate horizontal ocular deviation to one side indicates damage to the frontal lobe on the same side or less commonly the pons on the opposite side. This phenomenon is summarized by the following maxim: *The eyes look toward a hemispherical lesion and away from a brainstem lesion*. Seizures involving the frontal lobe drive the eyes to the opposite side, simulating a pontine destructive lesion. The eyes may occasionally turn paradoxically away from the side of a deep hemispherical lesion (“wrong-way eyes”). The eyes turn down and inward with thalamic and upper midbrain lesions, typically thalamic hemorrhage. “Ocular bobbing” describes brisk downward and slow upward movements of the eyes associated with loss of horizontal eye movements and is

diagnostic of bilateral pontine damage, usually from thrombosis of the basilar artery. "Ocular dipping" is a slower, arrhythmic downward movement followed by a faster upward movement in patients with normal reflex horizontal gaze; it usually indicates diffuse cortical anoxic damage.

The oculocephalic reflexes, elicited by moving the head from side to side or vertically and observing eye movements in the direction opposite to the head movement, depend on the integrity of the ocular motor nuclei and their interconnecting tracts that extend from the midbrain to the pons and medulla (Fig. 28-3). The movements, called somewhat inaccurately "doll's eyes," are normally suppressed in the awake patient with intact frontal lobes. The ability to elicit them therefore reflects both reduced cortical influence on the brainstem and intact brainstem pathways. The opposite, an absence of reflex eye movements, usually signifies damage within the brainstem but can result from overdoses of certain drugs. In this circumstance, normal pupillary size and light reaction distinguishes most drug-induced comas from structural brainstem damage. Oculocephalic maneuvers should not be attempted in patients with neck trauma, as vigorous head movements can precipitate or worsen a spinal cord injury.

Thermal, or "caloric," stimulation of the vestibular apparatus (oculovestibular response) provides a more intense stimulus for the oculocephalic reflex but provides essentially the same information. The test is performed by irrigating the external auditory canal with cold water in order to induce convection currents in the labyrinths. After a brief latency, the result is tonic deviation of both eyes to the side of cold-water irrigation. In comatose patients, nystagmus in the opposite direction may not occur. The acronym "COWS" has been used to remind generations of medical students of the direction of nystagmus—cold water opposite, warm water same—but since nystagmus is often absent in the opposite direction due to frontal lobe dysfunction in coma, this mnemonic does not often hold true.

The corneal reflex, elicited by touching the cornea with a wisp of cotton and observing bilateral lid closure, depends on the integrity of pontine pathways between the fifth (afferent) and both seventh (efferent) cranial nerves; it is a useful test of pontine function. CNS-depressant drugs diminish or eliminate the corneal responses soon after reflex eye movements are paralyzed but before the pupils become unreactive to light. The corneal response may be lost for a time on the side of an acute hemiplegia.

Respiratory Patterns These are of less localizing value in comparison to other brainstem signs. Shallow, slow, but regular breathing suggests metabolic or drug-induced depression of the medullary respiratory centers. Cheyne-Stokes respiration in its typical cyclic form, ending with a brief apneic period, signifies bihemispherical damage or metabolic suppression and commonly accompanies light coma. Rapid, deep (Kussmaul) breathing usually implies metabolic acidosis but may also occur with pontomesencephalic lesions. Agonal gasps are the result of lower brainstem (medullary) damage and are recognized as the terminal respiratory pattern of severe brain damage. Other cyclic breathing patterns have been described but are of lesser significance.

■ LABORATORY STUDIES AND IMAGING

The studies that are most useful in the diagnosis of coma are chemical-toxicologic analysis of blood and urine, cranial CT or MRI, EEG, and cerebrospinal fluid (CSF) examination. Arterial blood gas analysis is helpful in patients with lung disease and acid-base disorders. The metabolic aberrations commonly encountered in clinical practice are usually revealed by measurement of electrolytes, glucose, calcium, magnesium, osmolarity, and renal (blood urea nitrogen) and hepatic (NH_3) function. Toxicologic analysis may be necessary in cases of acute coma, when the diagnosis is not immediately clear. However, the presence of exogenous drugs or toxins, especially alcohol, does not exclude the possibility that other factors, particularly head trauma, are contributing to the clinical state. An ethanol level of 43 mmol/L

(0.2 g/dL) in nonhabituated patients generally causes impaired mental activity; a level of >65 mmol/L (0.3 g/dL) is associated with stupor. The development of tolerance may allow some chronic alcoholics to remain awake at levels >87 mmol/L (0.4 g/dL).

The availability of cranial CT and MRI has focused attention on causes of coma that are detectable by imaging (e.g., hemorrhage, tumor, or hydrocephalus). Resorting primarily to this approach, although at times expedient, is imprudent because most cases of coma (and confusion) are metabolic or toxic in origin. Furthermore, a normal CT scan does not exclude an anatomic lesion as the cause of coma; for example, early bilateral hemisphere infarction, acute brainstem infarction, encephalitis, meningitis, mechanical shearing of axons as a result of closed head trauma, sagittal sinus thrombosis, hypoxic injury, and subdural hematoma isodense to adjacent brain are some of the disorders that may not be detected. Sometimes imaging results can be misleading such as when small subdural hematomas or old strokes are found, but the patient's coma is due to intoxication. Additional imaging with CT angiography or MRI can be obtained if acute posterior circulation stroke is considered.

The EEG (Chap. 425) provides clues in metabolic or drug-induced states but is rarely diagnostic in these disorders. However, it is the essential test to reveal coma due to nonconvulsive seizures and shows fairly characteristic patterns in herpesvirus encephalitis and prion disease. The EEG may be further helpful in disclosing generalized slowing of the background activity, a reflection of the severity of an encephalopathy. Predominant high-voltage slowing (δ or triphasic waves) in the frontal regions is typical of metabolic coma, as from hepatic failure, and widespread fast (β) activity implicates overdose with sedative drugs (e.g., benzodiazepines). A special pattern of "alpha coma," defined by widespread, variable 8- to 12-Hz activity, superficially resembles the normal α rhythm of waking but, unlike normal α activity, is not altered by environmental stimuli. Alpha coma results from pontine or diffuse cortical damage and is associated with a poor prognosis. A unique EEG pattern in adults of "extreme delta brush" is characteristic of a specific (anti-*N*-methyl-*D*-aspartate [NMDA] receptor) form of autoimmune encephalitis. Normal α activity on the EEG, which is suppressed by stimulating the patient, also alerts the clinician to the locked-in syndrome, hysteria, or catatonia.

Lumbar puncture should be performed if no cause is readily apparent, as examination of the CSF remains indispensable in the diagnosis of various forms of meningitis and encephalitis. An imaging study should be performed prior to lumbar puncture to exclude a large intracranial mass lesion, which could lead to herniation with lumbar puncture. Blood cultures and administration of antibiotics should precede the imaging study if infectious meningitis is suspected (Chap. 138).

■ DIFFERENTIAL DIAGNOSIS OF COMA

(Table 28-1) The causes of coma can be divided into three broad categories: those without focal neurologic signs (e.g., metabolic and toxic encephalopathies); those with prominent focal signs (e.g., stroke, cerebral hemorrhage); and meningitis syndromes, characterized by fever or stiff neck and an excess of cells in the spinal fluid (e.g., bacterial meningitis, subarachnoid hemorrhage, encephalitis). Causes of sudden coma include drug ingestion, cerebral hemorrhage, trauma, cardiac arrest, epilepsy, and basilar artery occlusion. Coma that appears subacutely is usually related to a preexisting medical or neurologic problem or, less often, to secondary brain swelling surrounding a mass such as tumor or cerebral infarction.

The diagnosis of coma due to cerebrovascular disease can be difficult (Chap. 426). The most common diseases in this category are (1) basal ganglia and thalamic hemorrhage (acute but not instantaneous onset, vomiting, headache, hemiplegia, and characteristic eye signs); (2) pontine hemorrhage (sudden onset, pinpoint pupils, loss of reflex eye movements and corneal responses, ocular bobbing, posturing, and hyperventilation); (3) cerebellar hemorrhage (occipital headache, vomiting, gaze paresis, and inability to stand and walk); (4) basilar artery thrombosis (neurologic prodrome or transient ischemic attack warning spells, diplopia, dysarthria, vomiting, eye movement and corneal response abnormalities, and asymmetric limb paresis); and

TABLE 28-1 Differential Diagnosis of Coma

1. Diseases that cause no focal brainstem or lateralizing neurologic signs (CT scan is often normal)
 - a. Intoxications: alcohol, sedative drugs, opiates, etc.
 - b. Metabolic disturbances: anoxia, hyponatremia, hypernatremia, hypercalcemia, diabetic acidosis, nonketotic hyperosmolar hyperglycemia, hypoglycemia, uremia, hepatic coma, hypercarbia, Addisonian crisis, hypo- and hyperthyroid states, profound nutritional deficiency
 - c. Severe systemic infections: pneumonia, septicemia, typhoid fever, malaria, Waterhouse-Friderichsen syndrome
 - d. Shock from any cause
 - e. Status epilepticus, nonconvulsive status epilepticus, postictal states
 - f. Hyperperfusion syndromes including hypertensive encephalopathy, eclampsia, posterior reversible encephalopathy syndrome (PRES)
 - g. Severe hyperthermia, hypothermia
 - h. Concussion
 - i. Acute hydrocephalus
2. Diseases that cause focal brainstem or lateralizing cerebral signs (CT scan is typically abnormal)
 - a. Hemispherical hemorrhage (basal ganglionic, thalamic) or infarction (large middle cerebral artery territory) with secondary brainstem compression
 - b. Brainstem infarction due to basilar artery thrombosis or embolism
 - c. Brain abscess, subdural empyema
 - d. Epidural and subdural hemorrhage, brain contusion
 - e. Brain tumor with surrounding edema
 - f. Cerebellar and pontine hemorrhage and infarction
 - g. Widespread traumatic brain injury
 - h. Metabolic coma (see above) in the setting of preexisting focal damage
3. Diseases that cause meningeal irritation with or without fever, and with an excess of white blood cells or red blood cells in the CSF
 - a. Subarachnoid hemorrhage from ruptured aneurysm, arteriovenous malformation, trauma
 - b. Infectious meningitis and meningoencephalitis
 - c. Paraneoplastic and autoimmune encephalitis
 - d. Carcinomatous and lymphomatous meningitis

(5) subarachnoid hemorrhage (precipitous coma after sudden severe headache and vomiting). The most common stroke, infarction in the territory of the middle cerebral artery, does not cause coma, but edema surrounding large infarctions may expand over several days and cause coma from mass effect.

The syndrome of acute hydrocephalus accompanies many intracranial diseases, particularly subarachnoid hemorrhage. It is characterized by headache and sometimes vomiting that may progress quickly to coma with extensor posturing of the limbs, bilateral Babinski signs, small unreactive pupils, and impaired oculocephalic movements in the vertical direction. At times, the coma may be featureless without lateralizing signs, although papilledema is often present.

BRAIN DEATH

Brain death is a state of irreversible cessation of all cerebral and brainstem function with preservation of cardiac activity and maintenance of respiratory and somatic function by artificial means. It is the only type of brain damage recognized as morally, ethically, and legally equivalent to death. Criteria have been advanced for the diagnosis of brain death, and it is essential to adhere to consensus standards as multiple studies have shown variability in local practice. Given the implications of the diagnosis, clinicians must be thorough and precise in determining brain death. It is advisable to delay clinical testing for at least 24 h if a cardiac arrest has caused brain death or if the inciting disease is not known. Some centers advocate a brief period of observation between two examiners' tests during which the clinical signs of brain death are sustained.

Established criteria contain two essential elements, after assuring that no confounding factors (e.g., hypothermia, drug intoxication) are present: (1) widespread cortical destruction that is reflected by deep coma and unresponsiveness to all forms of stimulation; and (2) global

brainstem damage as demonstrated by absent pupillary light reaction, absent corneal reflexes, loss of oculovestibular reflexes, and destruction of the medulla, manifested by complete and irreversible apnea. Diabetes insipidus is often present but may only develop hours or days after the other clinical signs of brain death appear. The pupils are usually mid-sized but may be enlarged. Loss of deep tendon reflexes is not required because the spinal cord remains functional. Occasionally, other reflexes that originate from the spine may be present and should not preclude a diagnosis of brain death.

Demonstration that apnea is due to medullary damage requires that the PCO_2 be high enough to stimulate respiration during a test of spontaneous breathing. Apnea testing can be done by the use of preoxygenation with 100% oxygen prior to and following removal of the ventilator. CO_2 tension increases $\sim 0.3\text{--}0.4 \text{ kPa/min}$ ($2\text{--}3 \text{ mmHg/min}$) during apnea. Apnea is confirmed if no respiratory effort has been observed in the presence of a sufficiently elevated PCO_2 . The apnea test is usually stopped if there is cardiovascular instability and alternative means of testing can be employed.

An isoelectric EEG may be used as an optional confirmatory test for total cerebral damage. Radionuclide brain scanning, cerebral angiography, or transcranial Doppler measurements may be used to demonstrate the absence of blood flow when a confirmatory study is desired.

It is largely accepted in Western society that the ventilator can be disconnected from a brain-dead patient and that organ donation is subsequently possible. Good communication between the physician and the family is important with appropriate preparation of the family for brain death testing and diagnosis.

TREATMENT

Coma

The immediate goal in a comatose patient is prevention of further nervous system damage. Hypotension, hypoglycemia, hypercalcemia, hypoxia, hypercapnia, and hyperthermia should be corrected rapidly. Hyponatremia should be corrected slowly to avoid injury from osmotic demyelination (Chap. 307). An oropharyngeal airway is adequate to keep the pharynx open in a drowsy patient who is breathing normally. Tracheal intubation is indicated if there is apnea, upper airway obstruction, hypoventilation, or emesis, or if the patient is at risk for aspiration. Mechanical ventilation is required if there is hypoventilation or a need to induce hypocapnia in order to lower ICP. **The management of raised ICP is discussed in Chap. 307.** IV access is established and naloxone and dextrose are administered if opioid overdose or hypoglycemia are possibilities; thiamine is given along with glucose to avoid provoking Wernicke's encephalopathy in malnourished patients. In cases of suspected ischemic stroke including basilar thrombosis with brainstem ischemia, IV tissue plasminogen activator or mechanical embolectomy is often used after cerebral hemorrhage has been excluded and when the patient presents within established time windows for these interventions (Chap. 427). Physostigmine may awaken patients with anticholinergic-type drug overdose but should be used only with careful monitoring; many physicians believe that it should only be used to treat anticholinergic overdose-associated cardiac arrhythmias. The use of benzodiazepine antagonists offers some prospect of improvement after overdose; however, these drugs are not commonly used empirically in part due to their tendency to provoke seizures. Certain other toxic and drug-induced comas have specific treatments such as fomepizole for ethylene glycol ingestion.

Administration of hypotonic IV solutions should be monitored carefully in any serious acute brain illness because of the potential for exacerbating brain swelling. Cervical spine injuries must not be overlooked, particularly before attempting intubation or evaluation of oculocephalic responses. Fever and meningismus indicate an urgent need for examination of the CSF to diagnose meningitis. Whenever acute bacterial meningitis is suspected, antibiotics including at least vancomycin and a third-generation cephalosporin are typically administered rapidly along with dexamethasone (see Chap. 138).

■ PROGNOSIS

Some patients, especially children and young adults, may have ominous early clinical findings such as abnormal brainstem reflexes and yet recover; early prognostication outside of brain death therefore is unwise. Metabolic comas have a far better prognosis than traumatic ones. Systems for estimating prognosis in adults should be taken as approximations, and medical judgments must be tempered by factors such as age, underlying systemic disease, and general medical condition. In an attempt to collect prognostic information from large numbers of patients with head injury, the Glasgow Coma Scale was devised; it has predictive value in cases of brain trauma (see Chap. 443). For anoxic coma, clinical signs such as the pupillary and motor responses after 1 day, 3 days, and 1 week have predictive value; however, some prediction rules are less reliable in the setting of therapeutic hypothermia, and therefore, serial examinations and multimodal prognostication approaches are advised in this setting. For example, the absence of the cortical responses of the somatosensory evoked potentials has been shown to be a strong indicator of poor outcome following hypoxic injury.

The poor outcome of persistent vegetative and minimally conscious states has already been mentioned, but reports of a small number of patients displaying cortical activation on functional MRI in response to salient stimuli have begun to alter the perception of such individuals. In one series, about 10% of vegetative patients (mainly following traumatic brain injury) could activate their frontal or temporal lobes in response to requests by an examiner to imagine certain visuospatial tasks. Another series demonstrated that up to 15% of patients with various forms of acute brain injury and absence of behavioral responses to motor commands showed EEG activation in response to these commands. It is prudent to avoid generalizations from these findings, but the need for future studies of novel techniques to help communication and possibly recovery is needed.

■ FURTHER READING

- CLAASEN J et al: Detection of brain activation in unresponsive patients with acute brain injury. *N Engl J Med* 380:2497, 2019.
- EDLOW JA et al: Diagnosis of reversible causes of coma. *Lancet* 384:2064, 2014.
- GREER DM et al: Determination of brain death/death by neurologic criteria: The World Brain Death Project. *JAMA* 324:1078, 2020.
- MONTI MM et al: Willful modulation of brain activity in disorders of consciousness. *N Engl J Med* 362:579, 2010.
- POSNER JB et al: *Plum and Posner's Diagnosis of Stupor and Coma*, 5th ed. New York, Oxford University Press, 2019.
- WIJDICKS EFM: Predicting the outcome of a comatose patient at the bedside. *Pract Neurol* 20:26, 2020.

Neuropsychiatric and social deficits also arise in many dementia syndromes, manifesting as depression, apathy, anxiety, hallucinations, delusions, agitation, insomnia, sleep disturbances, compulsions, or disinhibition. The clinical course may be slowly progressive, as in Alzheimer's disease (AD); static, as in anoxic encephalopathy; or may fluctuate from day to day or minute to minute, as in dementia with Lewy bodies (DLB). Most patients with AD, the most prevalent form of dementia, begin with episodic memory impairment, but in other dementias, such as frontotemporal dementia (FTD), memory loss is not typically a presenting feature. **Focal cerebral disorders are discussed in Chap. 30 and illustrated in a video library in Chap. V2; detailed discussions of AD can be found in Chap. 431; FTD and related disorders in Chap. 432; vascular dementia in Chap. 433; DLB in Chap. 434; Huntington's disease (HD) in Chap. 436; and prion diseases in Chap. 438.**

FUNCTIONAL ANATOMY OF THE DEMENTIAS

Dementia syndromes result from the disruption of specific large-scale neuronal networks; the location and severity of synaptic and neuronal loss combine to produce the clinical features (Chap. 30). Behavior, mood, and attention are modulated by ascending noradrenergic, serotonergic, and dopaminergic pathways, whereas cholinergic signaling is critical for attention and memory functions. The dementias differ in the relative neurotransmitter deficit profiles; accordingly, accurate diagnosis guides effective pharmacologic therapy.

AD typically begins in the entorhinal region of the medial temporal lobe, spreads to the hippocampus and other limbic structures, and moves through the basal temporal areas and then into the lateral and posterior temporal and parietal neocortex, eventually causing a more widespread degeneration. Vascular dementia is associated with focal damage in a variable patchwork of cortical and subcortical regions or white matter tracts that disconnects nodes within distributed networks. In keeping with its anatomy, AD typically presents with episodic memory loss accompanied later by aphasia, executive dysfunction, or navigational problems. In contrast, dementias that begin in frontal or subcortical regions, such as FTD or HD, are less likely to begin with memory problems and more likely to present with difficulties with judgment, mood, executive control, movement, and behavior.

Lesions of frontal-striatal¹ pathways produce specific and predictable effects on behavior. The dorsolateral prefrontal cortex has connections with a central band of the caudate nucleus. Lesions of either the caudate or dorsolateral prefrontal cortex, or their connecting white matter pathways, may result in executive dysfunction, manifesting as poor organization and planning, decreased cognitive flexibility, and impaired working memory. The lateral orbital frontal cortex connects with the ventromedial caudate, and lesions of this system cause impulsiveness, distractibility, and disinhibition. The anterior cingulate cortex and adjacent medial prefrontal cortex project to the nucleus accumbens, and interruption of this system produces apathy, poverty of speech, emotional blunting, or even akinetic mutism. All corticostriatal systems also include topographically organized projections through the globus pallidus and thalamus, and damage to these nodes can likewise reproduce the clinical syndrome associated with the corresponding cortical or striatal injuries. Involvement of brainstem nuclei and cerebellar structures can further contribute to cognitive, behavioral, and motor manifestations.

■ THE CAUSES OF DEMENTIA

The single strongest risk factor for dementia is increasing age. The prevalence of disabling memory loss increases with each decade over age 50 and is usually associated with the microscopic changes of AD at autopsy. Yet some centenarians have intact memory function and no evidence of clinically significant dementia. Whether dementia is an inevitable consequence of normal human aging remains controversial although the prevalence increases with every decade of life.

29

Dementia

William W. Seeley, Gil D. Rabinovici,
Bruce L. Miller



Dementia, a syndrome with many causes, affects nearly 6 million people in the United States and results in a total annual health care cost in excess of \$300 billion. Dementia is defined as an acquired deterioration in cognitive abilities that impairs the successful performance of activities of daily living. Episodic memory, the ability to recall events specific in time and place, is the cognitive function most commonly lost; 10% of persons age >70 years and 20–40% of individuals age >85 years have clinically identifiable memory loss. In addition to memory, dementia may erode other mental faculties, including language, visuospatial, praxis, calculation, judgment, and problem-solving abilities.

¹The striatum comprises the caudate/putamen/nucleus accumbens.

TABLE 29-1 Differential Diagnosis of Dementia

Most Common Causes of Dementia	
Alzheimer's disease	Alcoholism ^a
Vascular dementia	PDD/LBD spectrum
Multi-infarct	Drug/medication intoxication ^a
Diffuse white matter disease (Binswanger's)	Limbic-predominant age-related TDP-43 encephalopathy
Less Common Causes of Dementia	
Vitamin deficiencies	Toxic disorders
Thiamine (B ₁): Wernicke's encephalopathy ^a	Drug, medication, and narcotic poisoning ^a
B ₁₂ (subacute combined degeneration) ^a	Heavy metal intoxication ^a
Nicotinic acid (pellagra) ^a	Organic toxins
Endocrine and other organ failure	Psychiatric
Hypothyroidism ^a	Depression (pseudodementia) ^a
Adrenal insufficiency and Cushing's syndrome ^a	Schizophrenia ^a
Hypo- and hyperparathyroidism ^a	Conversion disorder ^a
Renal failure ^a	Degenerative disorders
Liver failure ^a	Huntington's disease
Pulmonary failure ^a	Multisystem atrophy
Chronic infections	Hereditary ataxias (some forms)
HIV	Frontotemporal lobar degeneration spectrum
Neurosyphilis ^a	Multiple sclerosis
Papovavirus (JC virus) (progressive multifocal leukoencephalopathy)	Adult Down's syndrome with Alzheimer's disease
Tuberculosis, fungal, and protozoal ^a	ALS-parkinsonism-dementia complex of Guam
Whipple's disease ^a	Prion (Creutzfeldt-Jakob and Gerstmann-Sträussler-Scheinker diseases)
Head trauma and diffuse brain damage	Miscellaneous
Chronic traumatic encephalopathy	Sarcoidosis ^a
Chronic subdural hematoma ^a	Vasculitis ^a
Postanoxia	CADASIL, etc.
Postencephalitis	Acute intermittent porphyria ^a
Normal-pressure hydrocephalus ^a	Recurrent nonconvulsive seizures ^a
Intracranial hypotension	Additional conditions in children or adolescents
Neoplastic	Pantothenate kinase-associated neurodegeneration
Primary brain tumor ^a	Subacute sclerosing panencephalitis
Metastatic brain tumor ^a	Metabolic disorders (e.g., Wilson's and Leigh's diseases, leukodystrophies, lipid storage diseases, mitochondrial mutations)
Autoimmune (paraneoplastic) encephalitis ^a	

^aPotentially reversible dementia.

Abbreviations: ALS, amyotrophic lateral sclerosis; CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; LBD, Lewy body disease; PDD, Parkinson's disease dementia.

The many causes of dementia are listed in **Table 29-1**. The frequency of each condition depends on the age group under study, access of the group to medical care, country of origin, and perhaps racial or ethnic background. AD is the most common cause of dementia in Western countries, accounting for more than half of all patients. Vascular disease is the second most frequent cause for dementia and is particularly common in elderly patients or populations with limited access to medical care, where vascular risk factors are undertreated. Often, vascular brain injury is mixed with neurodegenerative disorders, particularly AD, making it difficult, even for the neuropathologist, to estimate the contribution of cerebrovascular disease to the cognitive disorder in an individual patient. Dementias associated with Parkinson's disease (PD) are common and may develop years after onset of a parkinsonian disorder, as seen with PD-related dementia (PDD), or they can occur concurrently with or preceding the motor syndrome, as in DLB.

Limbic-predominant aging-related TDP-43 encephalopathy (LATE) is common after age 70 and has been linked to declining episodic memory function. Chronic traumatic encephalopathy (CTE), a unique disease found in individuals with a history of repetitive head impacts (e.g., professional athletes in collision or fighting sports, military veterans exposed to multiple blasts), presents with changes in cognition, mood, behavior, or motor function. Mixed pathology is common, especially in older individuals. In patients under the age of 65, FTD rivals AD as the most common cause of dementia. Chronic intoxications, including those resulting from alcohol and prescription drugs, are an important and often treatable cause of dementia. Other disorders listed in Table 29-1 are uncommon but important because many are reversible. The classification of dementing illnesses into reversible and irreversible disorders is a useful approach to differential diagnosis. When effective treatments for the neurodegenerative conditions emerge, this dichotomy will become obsolete.

In a study of 1000 persons attending a memory disorders clinic, 19% had a potentially reversible cause of the cognitive impairment and 23% had a potentially reversible concomitant condition that may have contributed to the patient's impairment. The three most common potentially reversible diagnoses were depression, normal pressure hydrocephalus (NPH), and alcohol dependence; medication side effects are also common and should be considered in every patient (Table 29-1).

The term *rapidly progressive dementia* (RPD) is applied to illnesses that progress from initial symptom onset to dementia within a year or less; confusional states related to toxic/metabolic conditions are excluded. Although the prion proteinopathy Creutzfeldt-Jakob disease (CJD) ([Chap. 438](#)) is the classic cause of a rapidly progressive dementia, especially when associated with myoclonus, more often cases of RPD are due to AD or another neurodegenerative disorder, or to an autoimmune encephalitis.

Subtle cumulative decline in episodic memory is a common part of aging. This frustrating experience, often the source of jokes and humor, has historically been referred to as *benign forgetfulness of the elderly*. *Benign* means that it is not so progressive or serious that it impairs successful and productive daily functioning, although the distinction between benign and significant memory loss can be subtle. At age 85, the average person is able to learn and recall approximately one-half of the items (e.g., words on a list) that he or she could at age 18. The term *subjective cognitive decline* describes individuals who experience a subjective decline from their cognitive baseline but perform within normal limits for their age and educational attainment on formal neuropsychological testing. *Mild cognitive impairment* (MCI) is defined as a decline in cognition that is confirmed on objective cognitive testing but does not disrupt normal daily activities. MCI can be further subcategorized based on the presenting complaints and deficits (e.g., amnestic MCI, executive MCI). Factors that predict progression from MCI to an AD dementia include a prominent memory deficit, family history of dementia, presence of an apolipoprotein e4 (Apo e4) allele, small hippocampal volumes, an AD-like signature of cortical atrophy, low cerebrospinal fluid A β and elevated tau, or evidence of brain amyloid and tau deposition on positron emission tomography (PET) imaging.

The major degenerative dementias include AD, DBL, FTD and related disorders, HD, and prion diseases, including CJD. All are associated with the abnormal aggregation of a specific protein: A β ₄₂ and tau in AD; α -synuclein in DBL; tau, TAR DNA-binding protein of 43 kDa (TDP-43), or the FET family of proteins (*fused in sarcoma* [FUS], *Ewing sarcoma* [EWS], and *TBP-associated factor 15* [TAF15]) in FTD; huntingtin in HD; and misfolded prion protein (PrP^{Sc}) in CJD ([Table 29-2](#)).

The risk of developing dementia in late-life is associated with exposures and lifestyle factors that can operate across the life span. Modifiable risk factors include low education, hearing loss, traumatic brain injury, hypertension, diabetes mellitus, obesity, heavy alcohol use, smoking, depression, physical inactivity, and air pollution. Improved management of midlife vascular risk factors has been credited with a decreasing incidence of dementia observed in North America and Western Europe.

TABLE 29-2 The Molecular Basis for Degenerative Dementia

DEMENTIA	MOLECULAR BASIS	CAUSAL GENES (CHROMOSOME)	SUSCEPTIBILITY GENES	PATHOLOGIC FINDINGS
AD	A β /tau	APP (21), PS-1(14), PS-2(1) (<2% carry these mutations, most often in PS-1)	Apo e4(19)	Amyloid plaques, neurofibrillary tangles, and neuropil threads
FTD	Tau	MAPT exon and intron mutations (17) (about 10% of familial cases)	H1 MAPT haplotype	Tau neuronal and glial inclusions varying in morphology and distribution
	TDP-43	GRN (10% of familial cases), C9ORF72 (20%–30% of familial cases), rare VCP, very rare TARDBP, TBK1, TIA1		TDP-43 neuronal and glial inclusions varying in morphology and distribution
	FET	Very rare FUS		FET neuronal and glial inclusions varying in morphology and distribution
DLB	α -Synuclein	Very rare SNCA (4)	Unknown	α -Synuclein neuronal inclusions (Lewy bodies)
CJD	PrP ^{SC}	PRNP (20) (up to 15% of patients carry these dominant mutations)	Codon 129 homozygosity for methionine or valine	PrP ^{SC} deposition, panlaminar spongiosis

Abbreviations: AD, Alzheimer's disease; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FET, FUS/EWS/TAF-15; FTD, frontotemporal dementia.

APPROACH TO THE PATIENT

Dementias

Three major issues should be kept at the forefront: (1) What is the clinical diagnosis? (2) What component of the dementia syndrome is treatable or reversible? (3) Can the physician help to alleviate the burden on caregivers? A broad overview of the approach to dementia is shown in **Table 29-3**. The major degenerative dementias can usually be distinguished by the initial symptoms; neuropsychological, neuropsychiatric, and neurologic findings; and neuroimaging features (**Table 29-4**).

HISTORY

The history should concentrate on the onset, duration, and tempo of progression. An acute or subacute onset of confusion may be due to delirium (**Chap. 27**) and should trigger a search for intoxication, infection, or metabolic derangement. An elderly person with slowly progressive memory loss over several years is likely to suffer from AD. Nearly 75% of patients with AD begin with memory symptoms, but other early symptoms include anxiety or depression as well as difficulty managing money, driving, shopping, following instructions, finding words, or navigating. Personality change, disinhibition, and weight gain or compulsive eating suggest FTD, not AD. FTD is also suggested by prominent apathy, compulsion, loss of empathy for others, or progressive loss of speech fluency or single-word comprehension with relative sparing of memory and visuospatial abilities. The diagnosis of DLB is suggested by early visual hallucinations; parkinsonism; proneness to delirium or sensitivity to psychoactive medications; rapid eye movement (REM) behavior disorder (RBD; dramatic, sometimes violent, limb movements during dreaming [**Chap. 31**]); or Capgras syndrome, the delusion that a familiar person has been replaced by an impostor.

A history of stroke with irregular stepwise progression suggests vascular dementia. Vascular dementia is also commonly seen in the setting of hypertension, atrial fibrillation, peripheral vascular disease, smoking, and diabetes. In patients suffering from cerebrovascular disease, it can be difficult to determine whether the dementia is due to AD, vascular disease, or a mixture of the two because many of the risk factors for vascular dementia, including diabetes, high cholesterol, elevated homocysteine, and low exercise, are also risk factors for AD. Moreover, many patients with a major vascular contribution to their dementia lack a history of stepwise decline. Rapid progression with motor rigidity and myoclonus suggests CJD (**Chap. 438**). Seizures may indicate strokes or neoplasm but also occur in AD, particularly early-age-of-onset AD. Gait disturbance is common in vascular dementia, PD/DLB, or NPH. A history of high-risk sexual behaviors or intravenous drug use should trigger a search for central nervous system (CNS) infection, especially HIV or syphilis. A history of recurrent head trauma could indicate chronic

subdural hematoma, CTE, intracranial hypotension, or NPH. Subacute onset of severe amnesia and psychosis with mesial temporal T2/fluid-attenuated inversion recovery (FLAIR) hyperintensities on MRI should raise concern for autoimmune (paraneoplastic) encephalitis, sometimes in long-term smokers or other patients at risk for cancer. The spectrum of autoimmune etiologies producing

TABLE 29-3 Evaluation of the Patient with Dementia

ROUTINE EVALUATION	OPTIONAL FOCUSED TESTS	OCCASIONALLY HELPFUL TESTS
History	Psychometric testing	EEG
Physical examination	Chest x-ray	Parathyroid function
Laboratory tests	Lumbar puncture	Adrenal function
Thyroid function (TSH)	Liver function	Urine heavy metals
Vitamin B ₁₂	Renal function	RBC sedimentation rate
Complete blood count	Urine toxin screen	Angiogram
Electrolytes	HIV	Brain biopsy
CT/MRI	Apolipoprotein E	SPECT
	RPR or VDRL	PET
		Autoantibodies
Diagnostic Categories		
REVERSIBLE CAUSES	IRREVERSIBLE/DEGENERATIVE DEMENTIAS	PSYCHIATRIC DISORDERS
Examples	Examples	Depression
Hypothyroidism	Alzheimer's	Schizophrenia
Thiamine deficiency	Frontotemporal dementia	Conversion reaction
Vitamin B ₁₂ deficiency	Huntington's	
Normal pressure hydrocephalus	Dementia with Lewy bodies	
Subdural hematoma	Vascular	
Chronic infection	Leukoencephalopathies	
Brain tumor	Parkinson's	
Drug intoxication		
Autoimmune encephalopathy		
Associated Treatable Conditions		
	Depression	Agitation
	Seizures	Caregiver "burnout"
	Insomnia	Drug side effects

Abbreviations: CT, computed tomography; EEG, electroencephalogram; MRI, magnetic resonance imaging; PET, positron emission tomography; RBC, red blood cell; RPR, rapid plasma reagent (test); SPECT, single-photon emission computed tomography; TSH, thyroid-stimulating hormone; VDRL, venereal disease research laboratory (test for syphilis).

TABLE 29-4 Clinical Differentiation of the Major Dementias

DISEASE	FIRST SYMPTOM	MENTAL STATUS	NEUROPSYCHIATRY	NEUROLOGY	IMAGING
AD	Memory loss	Episodic memory loss	Irritability, anxiety, depression	Initially normal	Entorhinal cortex and hippocampal atrophy
FTD	Apathy, poor judgment/insight, speech/language, hyperorality	Frontal/executive and/or language; spares drawing	Apathy, disinhibition, overeating, compulsion	May have vertical gaze palsy, axial rigidity, dystonia, alien hand, or MND	Frontal, insular, and/or temporal atrophy; usually spares posterior parietal lobe
DLB	Visual hallucinations, REM sleep behavior disorder, delirium, Capgras syndrome, parkinsonism	Drawing and frontal/executive, spares memory, delirium-prone	Visual hallucinations, depression, sleep disorder, delusions	Parkinsonism	Posterior parietal atrophy, hippocampi larger than in AD
CJD	Dementia, mood, anxiety, movement disorders	Variable, frontal/executive, focal cortical, memory	Depression, anxiety, psychosis in some	Myoclonus, rigidity, parkinsonism	Cortical ribboning and basal ganglia or thalamus hyperintensity on diffusion/FLAIR MRI
Vascular	Often but not always sudden, variable, apathy, falls, focal weakness	Frontal/executive, cognitive slowing, can spare memory	Apathy, delusions, anxiety	Usually motor slowing, spasticity, can be normal	Cortical and/or subcortical infarctions, confluent white matter disease

Abbreviations: AD, Alzheimer's disease; CBD, cortical basal degeneration; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FLAIR, fluid-attenuated inversion recovery; FTD, frontotemporal dementia; MND, motor neuron disease; MRI, magnetic resonance imaging; REM, rapid eye movement.

RPD has rapidly expanded, and includes antibodies targeting leucine-rich glioma-inactivated 1 (LGI1; facioauricular dystonic seizures); contactin-associated protein-like 2 (Caspr2; insomnia, ataxia, myotonia); *N*-methyl-D-aspartate (NMDA)-receptor (psychosis, insomnia, dyskinesias); and α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA)-receptor (limbic encephalitis with relapses), among others (Chap. 94). Alcohol abuse creates risk for malnutrition and thiamine deficiency. Veganism, bowel irradiation, an autoimmune diathesis, a remote history of gastric surgery, and chronic therapy with histamine H₂-receptor antagonists for dyspepsia or gastroesophageal reflux predispose to B₁₂ deficiency. Certain occupations, such as working in a battery or chemical factory, might indicate heavy metal intoxication. Careful review of medication intake, especially for sedatives and analgesics, may raise the issue of chronic drug intoxication. An autosomal dominant family history is found in HD and in familial forms of AD, FTD, DLB, or prion disorders. A history of mood disorder, the recent death of a loved one, or depressive signs such as insomnia or weight loss, raise the possibility of depression-related cognitive impairment.

PHYSICAL AND NEUROLOGIC EXAMINATION

A thorough general and neurologic examination is essential to identify signs of nervous system involvement and search for clues suggesting a systemic disease that might be responsible for the cognitive disorder. Typical AD spares motor systems until late in the course. In contrast, patients with FTD often develop axial rigidity, supranuclear gaze palsy, or a motor neuron disease reminiscent of amyotrophic lateral sclerosis (ALS). In DLB, the initial symptoms may include a parkinsonian syndrome (resting tremor, cogwheel rigidity, bradykinesia, festinating gait), but DLB often starts with visual hallucinations or cognitive impairment, and symptoms referable to the lower brainstem (RBD, gastrointestinal, or autonomic problems) may arise years or even decades before parkinsonism or dementia. Corticobasal syndrome (CBS) features asymmetric akinesia and rigidity, dystonia, myoclonus, alien limb phenomena, pyramidal signs, and prefrontal deficits such as nonfluent aphasia with or without motor speech impairment, executive dysfunction, apraxia, or a behavioral disorder. Progressive supranuclear palsy (PSP) is associated with unexplained falls, axial rigidity, dysphagia, and vertical gaze deficits. CJD is suggested by the presence of diffuse rigidity, an akinetic mute state, and prominent, often startle-sensitive, myoclonus.

Hemiparesis or other focal neurologic deficits suggest vascular dementia or brain tumor. Dementia with a myelopathy and peripheral neuropathy suggests vitamin B₁₂ deficiency. Peripheral

neuropathy could also indicate another vitamin deficiency, heavy metal intoxication, thyroid dysfunction, Lyme disease, or vasculitis. Dry cool skin, hair loss, and bradycardia suggest hypothyroidism. Fluctuating confusion associated with repetitive stereotyped movements may indicate ongoing limbic, temporal, or frontal seizures. In the elderly, hearing impairment or visual loss may produce confusion and disorientation misinterpreted as dementia. Profound bilateral sensorineural hearing loss in a younger patient with short stature or myopathy, however, should raise concern for a mitochondrial disorder.

COGNITIVE AND NEUROPSYCHIATRIC EXAMINATION

Brief screening tools such as the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MOCA), the Tablet Based Cognitive Assessment Tool, and Cognistat can be used to capture dementia and follow progression. None of these tests is highly sensitive to early-stage dementia or reliably discriminates between dementia syndromes. The MMSE is a 30-point test of cognitive function, with each correct answer being scored as 1 point. It includes tests of: orientation (e.g., identify season/date/month/year/floor/hospital/town/state/country); registration (e.g., name and restate 3 objects); recall (e.g., remember the same three objects 5 minutes later); and language (e.g., name pencil and watch; repeat "no ifs ands or buts"; follow a 3-step command; obey a written command; and write a sentence and copy a design). In most patients with MCI and some with clinically apparent AD, bedside screening tests may be normal, and a more challenging and comprehensive set of neuropsychological tests will be required. When the etiology for the dementia syndrome remains in doubt, a specially tailored evaluation should be performed that includes tasks of working and episodic memory, executive function, language, and visuospatial and perceptual abilities. In AD, the early deficits involve episodic memory, category generation ("name as many animals as you can in 1 minute"), and visuoconstructive ability. Usually deficits in verbal or visual episodic memory are the first neuropsychological abnormalities detected, and tasks that require the patient to recall a long list of words or a series of pictures after a predetermined delay will demonstrate deficits in most patients. In FTD, the earliest deficits on cognitive testing involve executive control or language (speech or naming) functions, but some patients lack either finding despite profound social-emotional deficits. PDD or DLB patients have more severe deficits in executive and visuospatial function but do better on episodic memory tasks than patients with AD. Patients with vascular dementia often demonstrate a mixture of executive and visuospatial deficits, with prominent psychomotor slowing. In delirium, the most prominent deficits involve attention, working

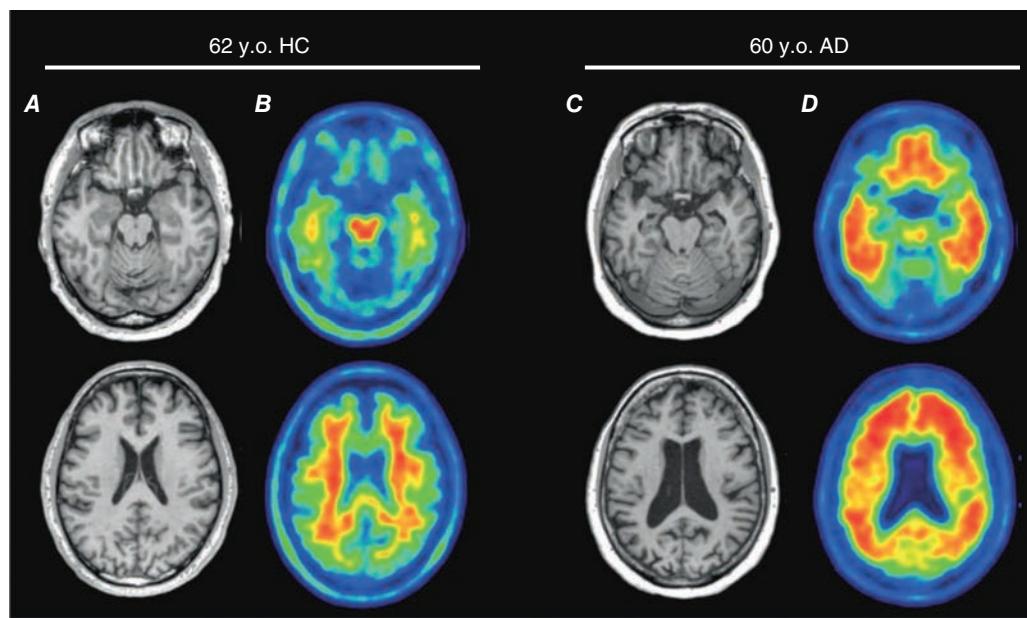


FIGURE 29-1 Alzheimer's disease (AD). Axial T1-weighted magnetic resonance images of a healthy 62-year-old (**A, B**) and a 60-year-old with AD (**C, D**). Note the diffuse atrophy, plus temporal lobe volume loss, in the patient with AD. A^β positron emission tomography (PET) with [¹¹C]PIB (**B** and **D**) reveals extensive radiotracer retention in neocortex bilaterally in AD, consistent with the known distribution of amyloid plaques. HC, healthy control. (Source: Gil Rabinovici, University of California, San Francisco and William Jagust, University of California, Berkeley.)

memory, and executive function, making the assessment of other cognitive domains challenging and often uninformative.

A functional assessment should also be performed to help the physician determine the day-to-day impact of the disorder on the patient's memory, community affairs, hobbies, judgment, dressing, and eating. Knowledge of the patient's functional abilities will help the clinician and the family to organize a therapeutic approach.

Neuropsychiatric assessment is important for diagnosis, prognosis, and treatment. In the early stages of AD, mild depressive features, social withdrawal, and irritability or anxiety are the most prominent psychiatric changes, but patients often maintain core social graces into the middle or late stages, when delusions, agitation, and sleep disturbance may emerge. In FTD, dramatic personality change with apathy, overeating, compulsions, disinhibition, and loss of empathy are early and common. DLB is associated with visual hallucinations, delusions related to person or place identity, RBD, and excessive daytime sleepiness. Dramatic fluctuations occur not only in cognition but also in arousal. Vascular dementia can present with psychiatric symptoms such as depression, anxiety, delusions, disinhibition, or apathy.

LABORATORY TESTS

The choice of laboratory tests in the evaluation of dementia is complex and should be tailored to the individual patient. The physician must take measures to avoid missing a reversible or treatable cause, yet no single treatable etiology is common; thus a screen must use multiple tests, each of which has a low yield. Cost/benefit ratios are difficult to assess, and many laboratory screening algorithms for dementia discourage multiple tests. Nevertheless, even a test with only a 1–2% positive rate is worth undertaking if the alternative is missing a treatable cause of dementia. Table 29-3 lists most screening tests for dementia. The American Academy of Neurology recommends the routine measurement of a complete blood count; electrolytes; glucose; renal, liver, and thyroid functions; a vitamin B₁₂ level; and a structural neuroimaging study (MRI or CT).

Neuroimaging studies, especially MRI, help to rule out primary and metastatic neoplasms, locate areas of infarction or inflammation, detect subdural hematomas, and suggest NPH or diffuse white matter disease. They also help to establish a regional pattern of atrophy. Support for the diagnosis of AD includes hippocampal

atrophy in addition to posterior-predominant cortical atrophy (Fig. 29-1). Focal frontal, insular, and/or anterior temporal atrophy suggests FTD (Chap. 432). DLB often features less prominent atrophy, with greater involvement of the amygdala than the hippocampus. In CJD, magnetic resonance (MR) diffusion-weighted imaging reveals restricted diffusion within the cortical ribbon and/or basal ganglia in most patients. Extensive multifocal white matter abnormalities suggest a vascular etiology (Fig. 29-2). Communicating hydrocephalus with vertex effacement (crowding of dorsal convexity gyri/sulci), gaping Sylvian fissures despite minimal cortical atrophy, and additional features shown in Fig. 29-3 suggest NPH. Single-photon emission computed tomography (SPECT) and fluoro-deoxyglucose PET scanning show temporal-parietal hypoperfusion or hypometabolism in AD and frontotemporal deficits in FTD, but abnormalities in these patterns can be detected with MRI alone in many patients. Recently, amyloid- and tau-PET imaging have shown promise for the diagnosis of AD. There are currently

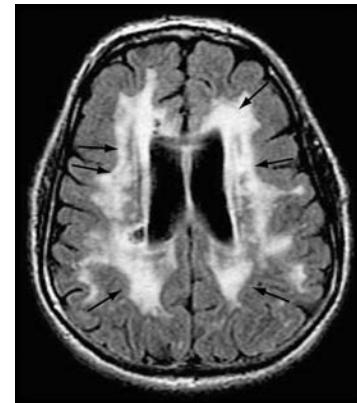


FIGURE 29-2 Diffuse white matter disease. Axial fluid-attenuated inversion recovery (FLAIR) magnetic resonance image through the lateral ventricles reveals multiple areas of hyperintensity (arrows) involving the periventricular white matter as well as the corona radiata and striatum. Although seen in some individuals with normal cognition, this appearance is more pronounced in patients with dementia of a vascular etiology.

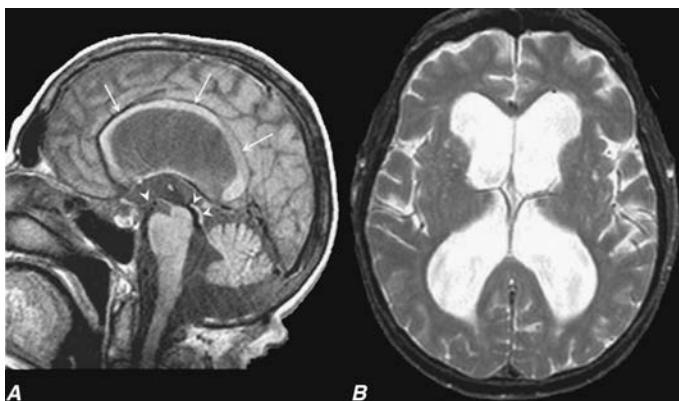


FIGURE 29-3 Normal pressure hydrocephalus. **A.** Sagittal T1-weighted MRI demonstrates dilation of the lateral ventricle and stretching of the corpus callosum (arrows), depression of the floor of the third ventricle (single arrowhead), and enlargement of the aqueduct (double arrowheads). Note the diffuse dilation of the lateral, third, and fourth ventricles with a patent aqueduct, typical of communicating hydrocephalus. **B.** Axial T2-weighted MRIs demonstrate dilation of the lateral ventricles. This patient underwent successful ventriculoperitoneal shunting.

three amyloid PET ligands (F18-florbetapir, F18-florbetaben, F18-flutemetamol) and one tau PET ligand (F18-flortaucipir) approved by the US Food and Drug Administration for clinical use. Amyloid PET ligands bind to diffuse and neuritic amyloid plaques, as well as to vascular amyloid deposits (prominent in cerebral amyloid angiopathy), while tau PET ligands bind to the paired helical filaments of tau characteristic of neurofibrillary tangles in AD (Chap. 431). Because amyloid plaques are also commonly found in cognitively normal older persons (~25% of individuals at age 65), the main clinical value of amyloid imaging is to exclude AD as the likely cause of dementia in patients who have negative scans. The spread of tau is more tightly linked to cognitive state (Chap. 431), and thus may be more useful than amyloid imaging for “ruling in” AD, as well as for disease staging. Once disease-modifying therapies become available, CSF or molecular PET biomarkers will likely be used to identify treatment candidates. In the meantime, the prognostic value of detecting brain amyloid in an asymptomatic elder to assess preclinical disease and risk of future cognitive decline remains a topic of vigorous investigation.

Lumbar puncture need not be done routinely in the evaluation of dementia, but it is indicated when CNS infection or inflammation are credible diagnostic possibilities. Cerebrospinal fluid (CSF) levels of $\text{A}\beta_{42}$ and tau proteins show differing patterns with the various dementias, and the presence of low $\text{A}\beta_{42}$ (or a low $\text{A}\beta_{42}/\text{A}\beta_{40}$ ratio), mild-moderately elevated CSF total tau, and elevated CSF phosphorylated tau (at residues 181 or 217) is highly suggestive of AD. Novel fully automated CSF A β and tau assays perform comparably to amyloid and tau PET respectively, though, as with PET, their routine use in the diagnosis of dementia is debated. Blood-based biomarkers for AD show promise as a less invasive screening tool but remain under development (Chap. 431). Formal psychometric testing helps to document the severity of cognitive disturbance, suggests psychogenic causes, and provides a more formal method for following the disease course. Electroencephalogram (EEG) is not routinely used but can help to suggest CJD (repetitive bursts of diffuse high-amplitude sharp waves, or “periodic complexes”) or an underlying nonconvulsive seizure disorder (epileptiform discharges). Brain biopsy (including meninges) is not advised except to diagnose vasculitis, neoplasms, or unusual infections when the diagnosis is uncertain. Systemic disorders with CNS manifestations, such as sarcoidosis, can often be confirmed through biopsy of lymph node or solid organ rather than brain. MR angiography should be considered when cerebral vasculitis or cerebral venous thrombosis is a possible cause of the dementia.

■ GLOBAL CONSIDERATIONS

Vascular dementia (Chap. 433) is more common in Asia due to the higher prevalence of intracranial atherosclerosis. Rates of vascular dementia are also on the rise in developing countries as vascular risk factors such as hypertension, hypercholesterolemia, and diabetes mellitus become more widespread. CNS infections, HIV (and associated opportunistic infections), syphilis, cysticercosis, and tuberculosis, likewise represent major contributors to dementia in the developing world. Systemic infection with SARS-CoV-2 may, in some individuals, have lasting effects on cognition due to involvement of brain microvasculature or to immunologically mediated white matter injury (acute disseminated encephalomyelitis [ADEM]) (Chap. 444). Some individuals complain of lasting fatigue, changes in mood, and cognitive difficulties, but the long-term prognosis for SARS-CoV-2-related cognitive impairment remains unknown. Isolated populations have also contributed to our understanding of neurodegenerative dementia. Kuru, the cannibalism-associated rapidly progressive dementia seen in tribal New Guinea, played a role in the discovery of human prion disease. Amyotrophic lateral sclerosis-parkinsonism-dementia complex of Guam (or, Lytico-bodig disease) is a poly-proteinopathy, often with tau, TDP-43, and alpha-synuclein aggregation. The root cause of the disease remains uncertain, but its incidence has declined sharply over the past 60 years.

TREATMENT

Dementia

The major goals of dementia management are to treat reversible causes and provide comfort and support to the patient and caregivers. Treatment of underlying causes includes thyroid replacement for hypothyroidism; vitamin therapy for thiamine or B₁₂ deficiency or for elevated serum homocysteine; antimicrobials for opportunistic infections or antiretrovirals for HIV; ventricular shunting for NPH; or surgical, radiation, and/or chemotherapeutic treatment for CNS neoplasms. Removal of cognition-impairing drugs or medications is essential when appropriate. If the patient’s cognitive complaints stem from a psychiatric disorder, vigorous treatment of the condition should be tried to eliminate the cognitive complaint or to confirm that it persists despite adequate resolution of the mood or anxiety symptoms. Patients with degenerative diseases may also be depressed or anxious, and those aspects of their condition often respond to therapy while not necessarily improving cognition. Antidepressants, such as selective serotonin reuptake inhibitors (SSRIs) or serotonin-norepinephrine reuptake inhibitors (SNRIs) (Chap. 452), which feature anxiolytic properties but few cognitive side effects, provide the mainstay of treatment when necessary. Anticonvulsants are used to control AD-associated seizures.

Agitation, hallucinations, delusions, and confusion are difficult to treat. These behavioral problems represent major causes for nursing home placement and institutionalization. Before treating these behaviors with medications, the clinician should aggressively seek out modifiable environmental or metabolic factors. Hunger, lack of exercise, toothache, constipation, urinary tract or respiratory infection, electrolyte imbalance, and drug toxicity all represent easily correctable causes that can be remedied without psychoactive drugs. Drugs such as phenothiazines and benzodiazepines may ameliorate the behavior problems but have untoward side effects such as sedation, rigidity, or dyskinesia; benzodiazepines can occasionally produce paradoxical disinhibition. Despite their unfavorable side effect profile, second-generation antipsychotics such as quetiapine (starting dose, 12.5–25 mg daily) can be used for patients with agitation, aggression, and psychosis, although the risk profile for these compounds is significant, including increased mortality in patients with dementia. When patients do not respond to treatment, it is usually a mistake to advance to higher doses or to use anticholinergic drugs (like diphenhydramine) or sedatives (such as barbiturates or benzodiazepines). It is important to recognize and treat depression; treatment can begin with a low dose of an

SSRI (e.g., escitalopram, starting dose 5 mg daily, target dose 5–10 mg daily) while monitoring for efficacy and toxicity. Sometimes apathy, visual hallucinations, depression, and other psychiatric symptoms respond to cholinesterase inhibitors, especially in DLB, obviating the need for other more toxic therapies.

Cholinesterase inhibitors are being used to treat AD (donepezil, rivastigmine, galantamine) and PDD (rivastigmine). Memantine is useful for some patients with moderate to severe AD; its major benefit relates to decreasing caregiver burden, most likely by decreasing resistance to dressing and grooming support. In moderate to severe AD, the combination of memantine and a cholinesterase inhibitor delayed nursing home placement in several studies, although other studies have not supported the efficacy of adding memantine to the regimen. Memantine should be used with great caution, or not at all, in patients with DLB, due to risk of worsening agitation and confusion. Therapies targeting the production, aggregation, and spread of misfolded proteins associated with dementia are under development. Recently the first drug in this class, the amyloid-beta targeting monoclonal antibody aducanumab, was approved by the United States Food & Drug Administration for treatment of Alzheimer's disease (Chap. 431). Other drugs under development target disease-associated neuroinflammation metabolic changes, synaptic loss, and neurotransmitter changes.

Proactive approaches reduce the occurrence of delirium in hospitalized patients. Frequent orientation, cognitive activities, sleep-enhancement measures, vision and hearing aids, and correction of dehydration are all valuable in decreasing the likelihood of delirium.

Nondrug behavior therapy has an important place in dementia management. The primary goals are to make the patient's life comfortable, uncomplicated, and safe. Preparing lists, schedules, calendars, and labels can be helpful in the early stages. It is also useful to stress familiar routines, walks, and simple physical exercises. For many demented patients, memory for events is worse than their ability to carry out routine activities, and they may still be able to take part in their favorite hobbies, sports, and social activities. Demented patients often object to losing control over familiar tasks such as driving, cooking, and handling finances. Attempts to help may be greeted with complaints, depression, or anger. Hostile responses on the part of the caregiver are counterproductive and sometimes even harmful. Reassurance, distraction, and calm positive statements are more productive when resistance is present. Eventually, tasks such as finances and driving must be assumed by others, and the patient will conform and adjust. Safety is an important issue that includes not only driving but controlling the kitchen, bathroom, and sleeping area environments, as well as stairways. These areas need to be monitored, supervised, and made as safe as possible. A move to a retirement complex, assisted-living center, or nursing home can initially increase confusion and agitation. Repeated reassurance, reorientation, and careful introduction to the new personnel will help to smooth the process. Providing activities that are known to be enjoyable to the patient can also help.

The clinician must pay special attention to frustration and depression among family members and caregivers. Caregiver guilt and burnout are common. Family members often feel overwhelmed and helpless and may vent their frustrations on the patient, each other, and health care providers. Caregivers should be encouraged to take advantage of day-care facilities and respite services. Education and counseling about dementia are important. Local and national support groups, such as the Alzheimer's Association (www.alz.org), can provide considerable help.

FURTHER READING

- BARTON C et al: Non-pharmacological management of behavioral symptoms in frontotemporal and other dementias. *Curr Neurol Neurosci Rep* 16:14, 2016.
- GRIEM J et al: Psychologic/functional forms of memory disorder. *Handb Clin Neurol* 139:407, 2017.
- WESLEY SF, FERGUSON D: Autoimmune encephalitides and rapidly progressive dementias. *Semin Neurol* 39:283, 2019.

30

Aphasia, Memory Loss, and Other Cognitive Disorders

M.-Marsel Mesulam

The cerebral cortex of the human brain contains ~20 billion neurons spread over an area of 2.5 m². The primary sensory and motor areas constitute 10% of the cerebral cortex. The rest is subsumed by modality-selective, heteromodal, paralimbic, and limbic areas collectively known as the *association cortex* (Fig. 30-1). The association cortex mediates the integrative processes that subserve cognition, emotion, and comportment. A systematic testing of these mental functions is necessary for the effective clinical assessment of the association cortex and its diseases. According to current thinking, there are no centers for "hearing words," "perceiving space," or "storing memories." Cognitive

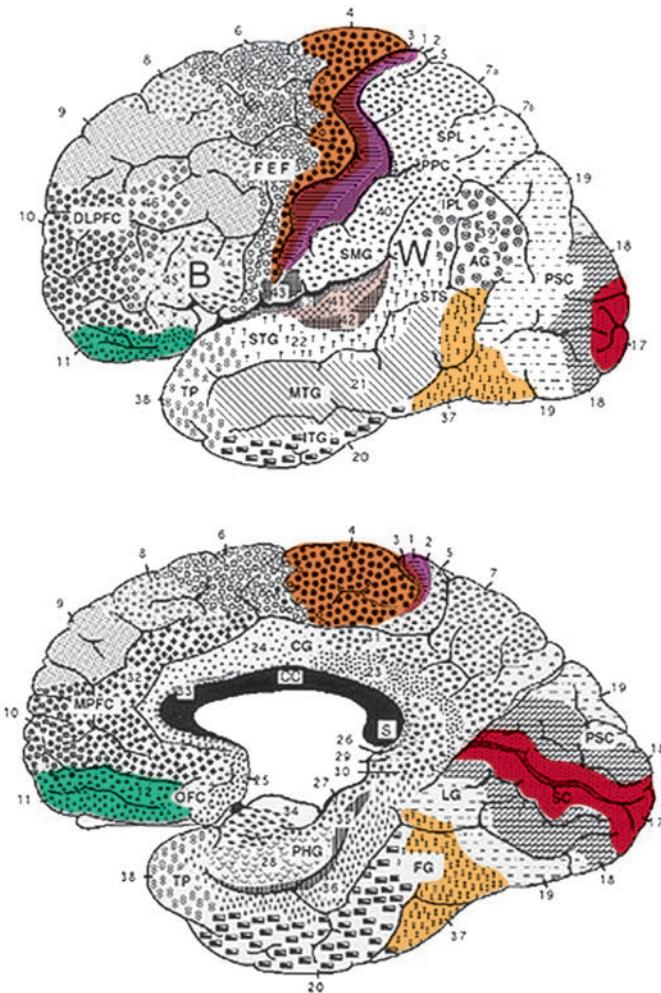


FIGURE 30-1 Lateral (top) and medial (bottom) views of the cerebral hemispheres. The numbers refer to the Brodmann cytoarchitectonic designations. Area 17 corresponds to the primary visual cortex, 41–42 to the primary auditory cortex, 1–3 to the primary somatosensory cortex, and 4 to the primary motor cortex. The rest of the cerebral cortex contains association areas. AG, angular gyrus; B, Broca's area; CC, corpus callosum; CG, cingulate gyrus; DLPFC, dorsolateral prefrontal cortex; FEF, frontal eye fields (premotor cortex); FG, fusiform gyrus; IPL, inferior parietal lobule; ITG, inferior temporal gyrus; LG, lingual gyrus; MPFC, medial prefrontal cortex; MTG, middle temporal gyrus; OFC, orbitofrontal cortex; PHG, parahippocampal gyrus; PPC, posterior parietal cortex; PSC, peristriate cortex; SC, striate cortex; SMG, supramarginal gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus; STS, superior temporal sulcus; TP, temporopolar cortex; W, Wernicke's area.

and behavioral functions (domains) are coordinated by intersecting *large-scale neural networks* that contain interconnected cortical and subcortical components. Five anatomically defined *large-scale networks* are most relevant to clinical practice: (1) a left-dominant perisylvian network for language, (2) a right-dominant parietofrontal network for spatial orientation, (3) an occipitotemporal network for face and object recognition, (4) a limbic network for episodic memory and emotional modulation, and (5) a prefrontal network for the executive control of cognition and comportment. Investigations based on functional imaging have also identified a *default mode network*, which becomes activated when the person is not engaged in a specific task requiring attention to external events. The clinical consequences of damage to this network are not yet fully defined.

THE LEFT PERISYLVIAN NETWORK FOR LANGUAGE AND APHASIAS

The production and comprehension of words and sentences is dependent on the integrity of a distributed network located along the perisylvian region of the language-dominant (usually left) hemisphere. One hub, situated in the inferior frontal gyrus, is known as *Broca's area*. Damage to this region impairs fluency of verbal output and the grammatical structure of sentences. The location of a second hub, critical for language comprehension, is less clearly settled. Accounts of patients with focal cerebrovascular lesions identified *Wernicke's area*, located at the parietotemporal junction, as a critical hub for word and sentence comprehension. Occlusive or embolic strokes involving this area interfere with the ability to understand spoken or written language as well as the ability to express thoughts through meaningful words and statements. However, investigations of patients with the neurodegenerative syndrome of primary progressive aphasia (PPA) have shown that sentence comprehension is a widely distributed faculty jointly subserved by Broca's and Wernicke's areas, and that the areas critical for word comprehension are more closely associated with the anterior temporal lobe than with Wernicke's area. All components of the language network are interconnected with each other and with surrounding parts of the frontal, parietal, and temporal lobes. Damage to this network gives rise to language impairments known as aphasia. Aphasia should be diagnosed only when there are deficits in the formal aspects of language, such as word finding, word choice, comprehension, spelling, or grammar. Dysarthria, apraxia of speech, and mutism do not by themselves lead to a diagnosis of aphasia. In ~90% of right-handers and 60% of left-handers, aphasia occurs only after lesions of the left hemisphere.

CLINICAL EXAMINATION

The clinical examination of language should include the assessment of naming, spontaneous speech, comprehension, repetition, reading, and writing. A deficit of naming (*anomia*) is the single most common finding in aphasic patients. When asked to name a common object, the patient may fail to come up with the appropriate word, may provide a circumlocutious description of the object ("the thing for writing"), or may come up with the wrong word (*paraphasia*). If the patient offers

an incorrect but related word ("pen" for "pencil"), the naming error is known as a *semantic paraphasia*; if the word approximates the correct answer but is phonetically inaccurate ("plentil" for "pencil"), it is known as a *phonemic paraphasia*. In most anomias, the patient cannot retrieve the appropriate name when shown an object but can point to the appropriate object when the name is provided by the examiner. This is known as a one-way (or retrieval-based) naming deficit. A two-way (comprehension-based or semantic) naming deficit exists if the patient can neither provide nor recognize the correct name. *Spontaneous speech* is described as "fluent" if it maintains appropriate output volume, phrase length, and melody or as "nonfluent" if it is sparse and halting and average utterance length is below four words. The examiner also should note the integrity of *grammar* as manifested by word order (syntax), tenses, suffixes, prefixes, plurals, and possessives. *Comprehension* can be tested by assessing the patient's ability to follow conversation, asking yes-no questions ("Can a dog fly?" "Does it snow in summer?"), asking the patient to point to appropriate objects ("Where is the source of illumination in this room?"), or asking for verbal definitions of single words. *Repetition* is assessed by asking the patient to repeat single words, short sentences, or strings of words such as "No ifs, ands, or buts." The testing of repetition with tongue twisters such as "hippopotamus" and "Irish constabulary" provides a better assessment of dysarthria and apraxia of speech than of aphasia. It is important to make sure that the number of words does not exceed the patient's attention span. Otherwise, the failure of repetition becomes a reflection of the narrowed attention span (auditory working memory) rather than an indication of an aphasic deficit caused by dysfunction of a hypothetical *phonological loop* in the language network. *Reading* should be assessed for deficits in reading aloud as well as comprehension. *Alexia* describes an inability to either read aloud or comprehend written words and sentences; *agraphia* (or *dysgraphia*) is used to describe an acquired deficit in spelling.

Aphasias can arise acutely in cerebrovascular accidents (CVAs) or gradually in neurodegenerative diseases. In CVAs, damage encompasses cerebral cortex as well as deep white matter pathways interconnecting otherwise unaffected cortical areas. The syndromes listed in **Table 30-1** are most applicable to this group, where gray matter and white matter at the lesion site are abruptly and jointly destroyed. Progressive neurodegenerative diseases can have cellular, laminar, and regional specificity for the cerebral cortex, giving rise to a different set of aphasias that will be described separately.

Wernicke's Aphasia Comprehension is impaired for spoken and written words and sentences. Language output is fluent but is highly paraphasic and circumlocutious. Paraphasic errors may lead to strings of neologisms, which lead to "jargon aphasia." Speech contains few substantive nouns. The output is therefore voluminous but uninformative. For example, a patient attempts to describe how his wife accidentally threw away something important, perhaps his dentures: "We don't need it anymore, she says. And with it when that was downstairs was my teeth-tick ... a ... den ... dentith ... my dentist. And they happened

TABLE 30-1 Clinical Features of Aphasias and Related Conditions Commonly Seen in Cerebrovascular Accidents

	COMPREHENSION	REPETITION OF SPOKEN LANGUAGE	NAMING	FLUENCY
Wernicke's	Impaired	Impaired	Impaired	Preserved or increased
Broca's	Preserved (except grammar)	Impaired	Impaired	Decreased
Global	Impaired	Impaired	Impaired	Decreased
Conduction	Preserved	Impaired	Impaired	Preserved
Nonfluent (anterior) transcortical	Preserved	Preserved	Impaired	Impaired
Fluent (posterior) transcortical	Impaired	Preserved	Impaired	Preserved
Isolation	Impaired	Echolalia	Impaired	No purposeful speech
Anomic	Preserved	Preserved	Impaired	Preserved except for word-finding pauses
Pure word deafness	Impaired only for spoken language	Impaired	Preserved	Preserved
Pure alexia	Impaired only for reading	Preserved	Preserved	Preserved

to be in that bag ... see? ... Where my two ... two little pieces of dentist that I use ... that I ... all gone. If she throws the whole thing away ... visit some friends of hers and she can't throw them away."

Gestures and pantomime do not improve communication. The patient may not realize that his or her language is incomprehensible and may appear angry and impatient when the examiner fails to decipher the meaning of a severely paraphasic statement. In some patients, this type of aphasia can be associated with severe agitation and paranoia. The ability to follow commands aimed at axial musculature may be preserved. The dissociation between the failure to understand simple questions ("What is your name?") in a patient who rapidly closes his or her eyes, sits up, or rolls over when asked to do so is characteristic of Wernicke's aphasia and helps differentiate it from deafness, psychiatric disease, or malingering. Patients with Wernicke's aphasia cannot express their thoughts in meaning-appropriate words and cannot decode the meaning of words in any modality of input. This aphasia therefore has expressive as well as receptive components. Repetition, naming, reading, and writing also are impaired.

The lesion site most commonly associated with Wernicke's aphasia caused by CVAs is the posterior portion of the language network. An embolus to the inferior division of the middle cerebral artery (MCA), to the posterior temporal or angular branches in particular, is the most common etiology (Chap. 426). Intracerebral hemorrhage, head trauma, and neoplasm are other causes of Wernicke's aphasia. A coexisting right hemianopia or superior quadrantanopia is common, and mild right nasolabial flattening may be found, but otherwise, the examination is often unrevealing. The paraphasic, neologistic speech in an agitated patient with an otherwise unremarkable neurologic examination may lead to the suspicion of a primary psychiatric disorder such as schizophrenia or mania, but the other components characteristic of acquired aphasia and the absence of prior psychiatric disease usually settle the issue. Prognosis for recovery of language function is guarded.

Broca's Aphasia Speech is nonfluent, labored, interrupted by many word-finding pauses, and usually dysarthric. It is impoverished in function words but enriched in meaning-appropriate nouns. Abnormal word order and the inappropriate deployment of *bound morphemes* (word endings used to denote tenses, possessives, or plurals) lead to a characteristic agrammatism. Speech is telegraphic and pithy but quite informative. In the following passage, a patient with Broca's aphasia describes his medical history: "I see ... the dotor, dotor sent me ... Bosson. Go to hospital. Dotor ... kept me beside. Two, tee days, doctor send me home."

Output may be reduced to a grunt or single word ("yes" or "no"), which is emitted with different intonations in an attempt to express approval or disapproval. In addition to fluency, naming and repetition are impaired. Comprehension of spoken language is intact except for syntactically difficult sentences with a passive voice structure or embedded clauses, indicating that Broca's aphasia is not just an "expressive" or "motor" disorder and that it also may involve a comprehension deficit in decoding syntax. Patients with Broca's aphasia can be tearful, easily frustrated, and profoundly depressed. Insight into their condition is preserved, in contrast to Wernicke's aphasia. Even when spontaneous speech is severely dysarthric, the patient may be able to display a relatively normal articulation of words when singing. This dissociation has been used to develop specific therapeutic approaches (melodic intonation therapy) for Broca's aphasia. Additional neurologic deficits include right facial weakness, hemiparesis or hemiplegia, and a buccofacial apraxia characterized by an inability to carry out motor commands involving oropharyngeal and facial musculature (e.g., patients are unable to demonstrate how to blow out a match or suck through a straw). The cause is most often infarction of Broca's area (the inferior frontal convolution; "B" in Fig. 30-1) and surrounding anterior perisylvian and insular cortex due to occlusion of the superior division of the MCA (Chap. 426). Mass lesions, including tumor, intracerebral hemorrhage, and abscess, also may be responsible. When the cause of Broca's aphasia is stroke, recovery of language function generally peaks within 2–6 months, after which time further progress is limited. Speech therapy is more successful than in Wernicke's aphasia.

Conduction Aphasia Speech output is fluent but contains many phonemic paraphasias, comprehension of spoken language is intact, and repetition is severely impaired. Naming elicits phonemic paraphasias, and spelling is impaired. Reading aloud is impaired, but reading comprehension is preserved. The responsible lesion, usually a CVA in the temporoparietal or dorsal perisylvian region, interferes with the function of the phonological loop interconnecting Broca's area with Wernicke's area. Occasionally, a transient Wernicke's aphasia may rapidly resolve into a conduction aphasia. The paraphasic and circumlocutious output in conduction aphasia interferes with the ability to express meaning, but this deficit is not nearly as severe as the one displayed by patients with Wernicke's aphasia. Associated neurologic signs in conduction aphasia vary according to the primary lesion site.

Transcortical Aphasias: Fluent and Nonfluent Clinical features of *fluent (posterior) transcortical aphasia* are similar to those of Wernicke's aphasia, but repetition is intact. The lesion site disconnects the intact core of the language network from other temporoparietal association areas. Associated neurologic findings may include hemianopia. Cerebrovascular lesions (e.g., infarctions in the posterior watershed zone) and neoplasms that involve the temporoparietal cortex posterior to Wernicke's area are common causes. The features of *nonfluent (anterior) transcortical aphasia* are similar to those of Broca's aphasia, but repetition is intact and agrammatism is less pronounced. The neurologic examination may be otherwise intact, but a right hemiparesis also can exist. The lesion site disconnects the intact language network from prefrontal areas of the brain and usually involves the anterior watershed zone between anterior and MCA territories or the supplementary motor cortex in the territory of the anterior cerebral artery.

Global and Isolation Aphasias *Global aphasia* represents the combined dysfunction of Broca's and Wernicke's areas and usually results from strokes that involve the entire MCA distribution in the left hemisphere. Speech output is nonfluent, and comprehension of language is severely impaired. Related signs include right hemiplegia, hemisensory loss, and homonymous hemianopia. *Isolation aphasia* represents a combination of the two transcortical aphasias. Comprehension is severely impaired, and there is no purposeful speech output. The patient may parrot fragments of heard conversations (*echolalia*), indicating that the neural mechanisms for repetition are at least partially intact. This condition represents the pathologic function of the language network when it is isolated from other regions of the brain. Broca's and Wernicke's areas tend to be spared, but there is damage to the surrounding frontal, parietal, and temporal cortex. Lesions are patchy and can be associated with anoxia, carbon monoxide poisoning, or complete watershed zone infarctions.

Anomic Aphasia This form of aphasia may be considered the "minimal dysfunction" syndrome of the language network. Articulation, comprehension, and repetition are intact, but confrontation naming, word finding, and spelling are impaired. Word-finding pauses are uncommon, so language output is fluent but paraphasic, circumlocutious, and uninformative. The lesion sites can be anywhere within the left hemisphere language network, including the middle and inferior temporal gyri. *Anomic aphasia* is the single most common language disturbance seen in head trauma, metabolic encephalopathy, and Alzheimer's disease.

Pure Word Deafness The most common causes are either bilateral or left-sided MCA strokes affecting the superior temporal gyrus. The net effect of the underlying lesion is to interrupt the flow of information from the auditory association cortex to the language network. Patients have no difficulty understanding written language and can express themselves well in spoken or written language. They have no difficulty interpreting and reacting to environmental sounds if the primary auditory cortex and auditory association areas of the right hemisphere are spared. Because auditory information cannot be conveyed to the language network, however, it cannot be decoded into neural word representations, and the patient reacts to speech as if it were in an alien tongue that cannot be deciphered. Patients cannot

repeat spoken language but have no difficulty naming objects. In time, patients with pure word deafness teach themselves lipreading and may appear to have improved. There may be no additional neurologic findings, but agitated paranoid reactions are common in the acute stages. Cerebrovascular lesions are the most common cause.

Pure Alexia Without Agraphia This is the visual equivalent of pure word deafness. The lesions (usually a combination of damage to the left occipital cortex and to a posterior sector of the corpus callosum—the splenium) interrupt the flow of visual input into the language network. There is usually a right hemianopia, but the core language network remains unaffected. The patient can understand and produce spoken language, name objects in the left visual hemifield, repeat, and write. However, the patient acts as if illiterate when asked to read even the simplest sentence because the visual information from the written words (presented to the intact left visual hemifield) cannot reach the language network. Objects in the left hemifield may be named accurately because they activate nonvisual associations in the right hemisphere, which in turn can access the language network through transcallosal pathways anterior to the splenium. Patients with this syndrome also may lose the ability to name colors, although they can match colors. This is known as a *color anomia*. The most common etiology of pure alexia is a vascular lesion in the territory of the posterior cerebral artery or an infiltrating neoplasm in the left occipital cortex that involves the optic radiations as well as the crossing fibers of the splenium. Because the posterior cerebral artery also supplies medial temporal components of the limbic system, a patient with pure alexia also may experience an amnesia, but this is usually transient because the limbic lesion is unilateral.

Apraxia and Aphemia *Apraxia* designates a complex motor deficit that cannot be attributed to pyramidal, extrapyramidal, cerebellar, or sensory dysfunction and that does not arise from the patient's failure to understand the nature of the task. *Apraxia of speech* is used to designate articulatory abnormalities in the duration, fluidity, and stress of syllables that make up words. It can arise with CVAs in the posterior part of Broca's area or in the course of frontotemporal lobar degeneration (FTLD) with tauopathy. *Aphemia* is a severe form of acute speech apraxia that presents with severely impaired fluency (often mutism). Recovery is the rule and involves an intermediate stage of hoarse whispering. Writing, reading, and comprehension are intact, and so this is not a true aphasic syndrome. CVAs in parts of Broca's area or subcortical lesions that undercut its connections with other parts of the brain may be present. Occasionally, the lesion site is on the medial aspects of the frontal lobes and may involve the supplementary motor cortex of the left hemisphere. *Ideomotor apraxia* is diagnosed when commands to perform a specific motor act ("cough," "blow out a match") or pantomime the use of a common tool (a comb, hammer, straw, or toothbrush) in the absence of the real object cannot be followed. The patient's ability to comprehend the command is ascertained by demonstrating multiple movements and establishing that the correct one can be recognized. Some patients with this type of apraxia can imitate the appropriate movement when it is demonstrated by the examiner and show no impairment when handed the real object, indicating that the sensorimotor mechanisms necessary for the movement are intact. Some forms of ideomotor apraxia represent a disconnection of the language network from pyramidal motor systems so that commands to execute complex movements are understood but cannot be conveyed to the appropriate motor areas. *Buccofacial apraxia* involves apraxic deficits in movements of the face and mouth. Ideomotor *limb apraxia* encompasses apraxic deficits in movements of the arms and legs. Ideomotor apraxia almost always is caused by lesions in the left hemisphere and is commonly associated with aphasic syndromes, especially Broca's aphasia and conduction aphasia. Because the handling of real objects is not impaired, ideomotor apraxia by itself causes no major limitation of daily living activities. Patients with lesions of the anterior corpus callosum can display ideomotor apraxia confined to the left side of the body, a sign known as *sympathetic dyspraxia*. A severe form of sympathetic dyspraxia, known as the *alien hand* syndrome, is

characterized by additional features of motor disinhibition on the left hand. *Ideational apraxia* refers to a deficit in the sequencing of goal-directed movements in patients who have no difficulty executing the individual components of the sequence. For example, when the patient is asked to pick up a pen and write, the sequence of uncapping the pen, placing the cap at the opposite end, turning the point toward the writing surface, and writing may be disrupted, and the patient may be seen trying to write with the wrong end of the pen or even with the removed cap. These motor sequencing problems usually are seen in the context of confusional states and dementias rather than focal lesions associated with aphasic conditions. *Limb-kinetic apraxia* involves clumsiness in the use of tools or objects that cannot be attributed to sensory, pyramidal, extrapyramidal, or cerebellar dysfunction. This condition can emerge in the context of focal premotor cortex lesions or *corticobasal degeneration* and can interfere with the use of tools and utensils.

Gerstmann's Syndrome The combination of *acalculia* (impairment of simple arithmetic), *dysgraphia* (impaired writing), *finger anomia* (an inability to name individual fingers such as the index and thumb), and *right-left confusion* (an inability to tell whether a hand, foot, or arm of the patient or examiner is on the right or left side of the body) is known as Gerstmann's syndrome. In making this diagnosis, it is important to establish that the finger and left-right naming deficits are not part of a more generalized anomia and that the patient is not otherwise aphasic. When Gerstmann's syndrome arises acutely and in isolation, it is commonly associated with damage to the inferior parietal lobule (especially the angular gyrus) in the left hemisphere.

Pragmatics and Prosody *Pragmatics* refers to aspects of language that communicate attitude, affect, and the figurative rather than literal aspects of a message (e.g., "green thumb" does not refer to the actual color of the finger). One component of pragmatics, *prosody*, refers to variations of melodic stress and intonation that influence attitude and the inferential aspect of verbal messages. For example, the two statements "He is clever." and "He is *clever?*" contain an identical word choice and syntax but convey vastly different messages because of differences in the intonation with which the statements are uttered. Damage to right hemisphere regions corresponding to Broca's area impairs the ability to introduce meaning-appropriate prosody into spoken language. The patient produces grammatically correct language with accurate word choice, but the statements are uttered in a monotone that interferes with the ability to convey the intended stress and effect. Patients with this type of *aprosodia* give the mistaken impression of being depressed or indifferent. Other aspects of pragmatics, especially the ability to infer the figurative aspect of a message, become impaired by damage to the right hemisphere or frontal lobes.

Subcortical Aphasia Damage to subcortical components of the language network (e.g., the striatum and thalamus of the left hemisphere) also can lead to aphasia. The resulting syndromes contain combinations of deficits in the various aspects of language but rarely fit the specific patterns described in Table 30-1. In a patient with a CVA, an anomic aphasia accompanied by dysarthria or a fluent aphasia with hemiparesis should raise the suspicion of a subcortical lesion site.

CLINICAL PRESENTATION AND DIAGNOSIS OF PPA Aphasias caused by CVAs start suddenly and display maximal deficits at the onset. These are the "classic" aphasias described above. Aphasias caused by neurodegenerative diseases have an insidious onset and relentless progression. The neuropathology can be selective not only for gray matter but also for specific layers and cell types. The clinico-anatomic patterns are therefore different from those described in Table 30-1.

Several neurodegenerative syndromes, such as typical Alzheimer-type (amnestic; **Chap. 431**) and frontotemporal (behavioral; **Chap. 432**) dementias, can also include language impairments as the disease progresses. In these cases, the aphasia is an ancillary component of the overall syndrome. A diagnosis of primary progressive aphasia (PPA) is justified only if the language disorder (i.e., aphasia) arises in relative isolation, becomes the primary concern that brings the patient to medical attention, and remains the most salient deficit for 1–2 years. PPA

can be caused by either FTLD or Alzheimer's disease (AD) pathology. Rarely, an identical syndrome can be caused by Creutzfeldt-Jacob disease (CJD) but with a more rapid progression (**Chap. 438**).

LANGUAGE IN PPA The impairments of language in PPA have slightly different patterns from those seen in CVA-caused aphasias. For example, the full syndrome of Wernicke's aphasia is almost never seen in PPA, confirming the view that sentence comprehension and word comprehension are controlled by different regions of the language network. Three major subtypes of PPA can be recognized.

Agrammatic PPA The *agrammatic variant* is characterized by consistently low fluency and impaired grammar but intact word comprehension. It most closely resembles Broca's aphasia or anterior transcortical aphasia but usually lacks the right hemiparesis or dysarthria and may have more profound impairments of grammar. Peak sites of neuronal loss (gray matter atrophy) include the left inferior frontal gyrus where Broca's area is located. The neuropathology is usually a FTLD with tauopathy but can also be an atypical form of AD pathology.

Semantic PPA The *semantic variant* is characterized by preserved fluency and syntax but poor single-word comprehension and profound two-way naming impairments. This kind of aphasia is not seen with CVAs. It differs from Wernicke's aphasia or posterior transcortical aphasia because speech is usually informative and repetition is intact. Comprehension of sentences is relatively preserved if the meaning is not too dependent on words that fail to be understood allowing the patient to surmise the gist of the conversation through contextual cues. Such patients may appear unimpaired in the course of casual small talk but become puzzled upon encountering an undecipherable word such as "pumpkin" or "umbrella." Peak atrophy sites are located in the left anterior temporal lobe, indicating that this part of the brain plays a critical role in the comprehension of words, especially words that denote concrete objects. This is a part of the brain that was not included within the classic language network, probably because it is not a common site for focal CVAs. The neuropathology is frequently an FTLD with abnormal precipitates of the 43-kDa transactive response DNA-binding protein TDP-43 of type C.

Logopenic PPA The *logopenic variant* is characterized by preserved syntax and comprehension but frequent and severe word-finding pauses, anomia, circumlocutions, and simplifications during spontaneous speech. Repetition is usually impaired. Peak atrophy sites are located in the temporoparietal junction and posterior temporal lobe, partially overlapping with traditional location of Wernicke's area. However, the comprehension impairment of *Wernicke's aphasia* is absent probably because the underlying deep white matter, frequently damaged by CVAs, remains relatively intact in PPA. The repetition impairment suggests that parts of Wernicke's area are critical for phonological loop functionality. In contrast to Broca's aphasia or agrammatic PPA, the interruption of fluency is variable so that speech may appear entirely normal if the patient is allowed to engage in small talk. Logopenic PPA resembles the anomic aphasia of Table 30-1 but usually has longer and more frequent word-finding pauses. When repetition is impaired, the aphasia resembles the *conduction aphasia* in Table 30-1. Of all PPA subtypes, this is the one most commonly associated with the pathology of AD, but FTLD can also be the cause. In addition to these three major subtypes, there is also a *mixed* type of PPA where grammar, fluency, and word comprehension are jointly impaired. This is most like the global aphasia of Table 30-1. Rarely, PPA can present with patterns reminiscent of *pure word deafness* or *Gerstmann's syndrome*.

THE PARIETOFRONTAL NETWORK FOR SPATIAL ORIENTATION

Adaptive spatial orientation is subserved by a large-scale network containing three major cortical components. The *cingulate cortex* provides access to a motivational mapping of the extrapersonal space, the *posterior parietal cortex* to a sensorimotor representation of salient extrapersonal events, and the *frontal eye fields* to motor strategies for attentional

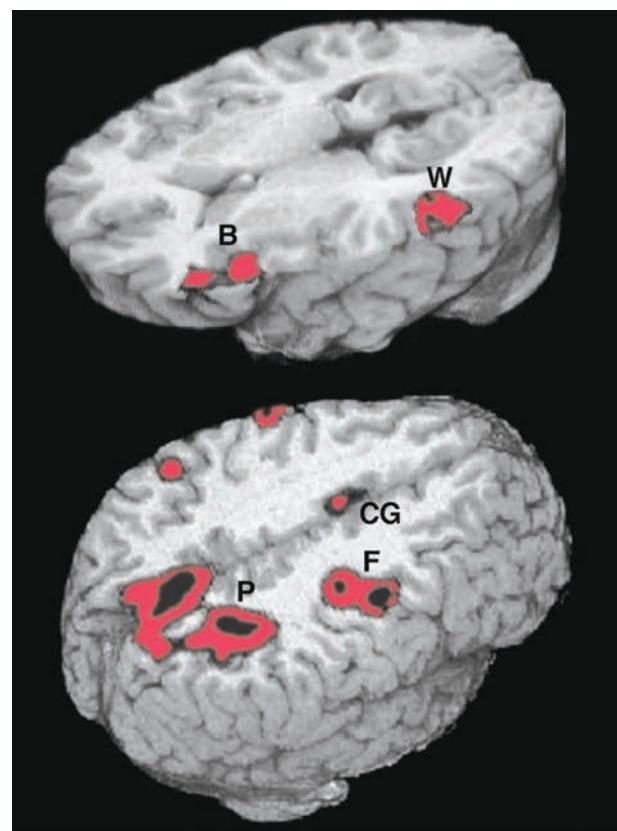


FIGURE 30-2 Functional magnetic resonance imaging of language and spatial attention in neurologically intact subjects. The red and black areas show regions of task-related significant activation. (Top) The subjects were asked to determine if two words were synonymous. This language task led to the simultaneous activation of the two components of the language network, Broca's area (B) and Wernicke's area (W). The activations are exclusively in the left hemisphere. (Bottom) The subjects were asked to shift spatial attention to a peripheral target. This task led to the simultaneous activation of the three epicenters of the attentional network: the posterior parietal cortex (P), the frontal eye fields (F), and the cingulate gyrus (CG). The activations are predominantly in the right hemisphere. (Courtesy of Darren Gitelman, MD.)

behaviors (Fig. 30-2). Subcortical components of this network include the striatum and the thalamus. Damage to this network can undermine the distribution of attention within the extrapersonal space, giving rise to hemispatial neglect, simultanagnosia, and object finding failures. The integration of egocentric (self-centered) with allocentric (object-centered) coordinates can also be disrupted, giving rise to impairments in route finding, the ability to avoid obstacles, and the ability to dress.

■ HEMISPATIAL NEGLECT

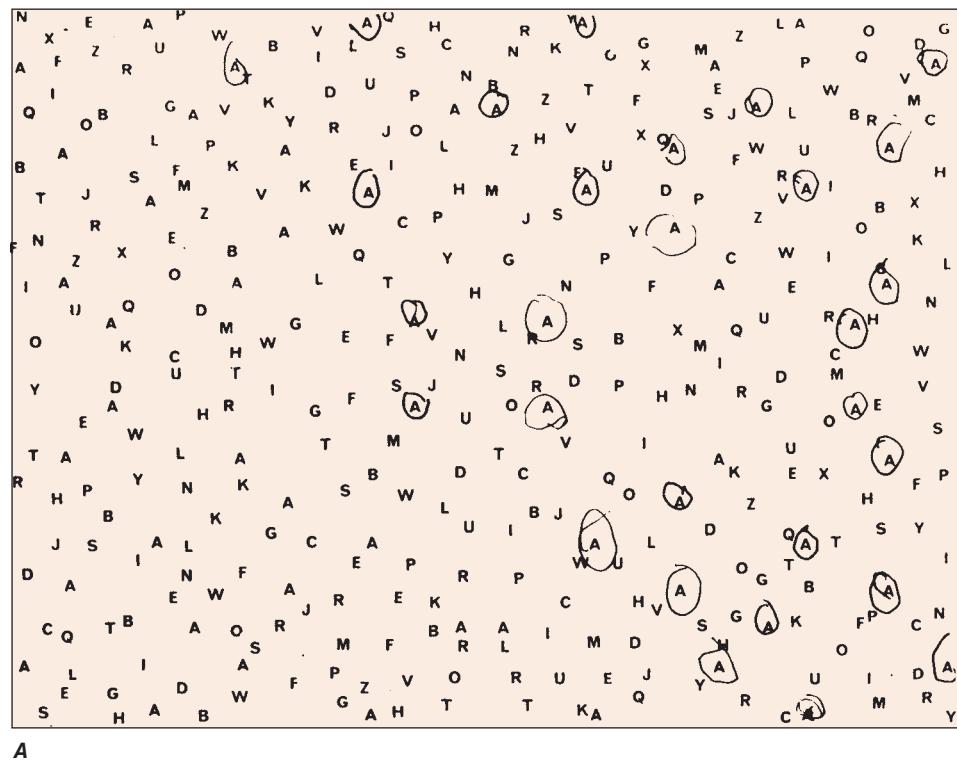
Contralesional hemispatial neglect represents one outcome of damage to the cortical or subcortical components of this network. *The traditional view that hemispatial neglect always denotes a parietal lobe lesion is inaccurate.* According to one model of spatial cognition, the right hemisphere directs attention within the *entire* extrapersonal space, whereas the left hemisphere directs attention mostly within the contralateral right hemispace. Consequently, left hemisphere lesions do not give rise to much contralesional neglect because the global attentional mechanisms of the right hemisphere can compensate for the loss of the *contralaterally* directed attentional functions of the left hemisphere. Right hemisphere lesions, however, give rise to severe contralesional left hemispatial neglect because the unaffected left hemisphere does not contain ipsilateral attentional mechanisms. This model is consistent with clinical experience, which shows that contralesional neglect is more common, more severe, and longer lasting after damage to the right hemisphere than after damage to the left hemisphere. Severe neglect for the right hemispace is rare, even in left-handers with left hemisphere lesions.

Clinical Examination Patients with severe neglect may fail to dress, shave, or groom the left side of the body; fail to eat food placed on the left side of the tray; and fail to read the left half of sentences. When asked to copy a simple line drawing, the patient fails to copy detail on the left, and when the patient is asked to write, there is a tendency to leave an unusually wide margin on the left. Two bedside tests that are useful in assessing neglect are *simultaneous bilateral stimulation* and *visual target cancellation*. In the former, the examiner provides either unilateral or simultaneous bilateral stimulation in the visual, auditory, and tactile modalities. After right hemisphere injury, patients who have no difficulty detecting unilateral stimuli on either side experience the bilaterally presented stimulus as coming only from the right. This phenomenon is known as *extinction* and is a manifestation of the sensory-representational aspect of hemispatial neglect. In the target detection task, targets (e.g., A's) are interspersed with foils (e.g., other letters of the alphabet) on a 21.5- to 28.0-cm (8.5–11 in.) sheet

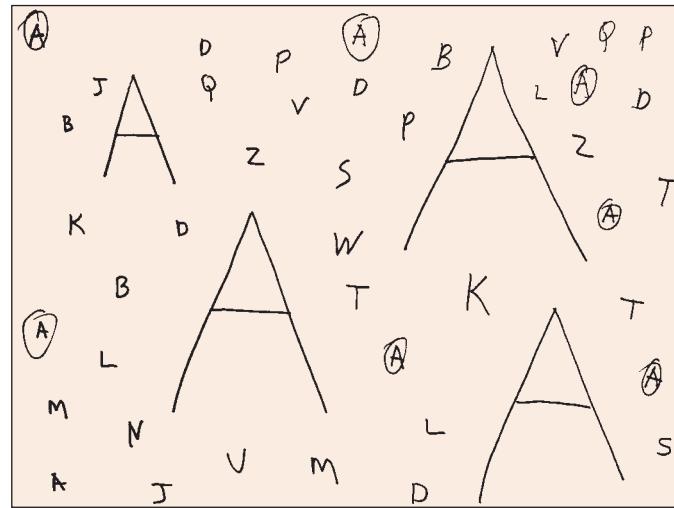
of paper, and the patient is asked to circle all the targets. A failure to detect targets on the left is a manifestation of the exploratory (motor) deficit in hemispatial neglect (Fig. 30-3A). Hemianopia is not by itself sufficient to cause the target detection failure because the patient is free to turn the head and eyes to the left. Target detection failures therefore reflect a distortion of spatial attention, not just of sensory input. Some patients with neglect also may deny the existence of hemiparesis and may even deny ownership of the paralyzed limb, a condition known as *anosognosia*.

BÁLINT'S SYNDROME, SIMULTANAGNOSIA, DRESSING APRAXIA, CONSTRUCTION APRAXIA, AND ROUTE-FINDING IMPAIRMENTS

Bilateral involvement of the network for spatial attention, especially its parietal components, leads to a state of severe spatial disorientation known as *Bálint's syndrome*. Bálint's syndrome involves deficits in the



A



B

FIGURE 30-3 **A.** A 47-year-old man with a large frontoparietal lesion in the right hemisphere was asked to circle all the A's. Only targets on the right are circled. This is a manifestation of left hemispatial neglect. **B.** A 70-year-old woman with a 2-year history of degenerative dementia was able to circle most of the small targets but ignored the larger ones. This is a manifestation of simultanagnosia.

orderly visuomotor scanning of the environment (*oculomotor apraxia*), accurate manual reaching toward visual targets (*optic ataxia*), and the ability to integrate visual information in the center of gaze with more peripheral information (*simultanagnosia*). A patient with simultanagnosia “misses the forest for the trees.” For example, a patient who is shown a table lamp and asked to name the object may look at its circular base and call it an ashtray. Some patients with simultanagnosia report that objects they look at may vanish suddenly, probably indicating an inability to compute the oculomotor return to the original point of gaze after brief saccadic displacements. Movement and distracting stimuli greatly exacerbate the difficulties of visual perception. Simultanagnosia can occur without the other two components of Bálint’s syndrome, especially in association with AD.

A modification of the letter cancellation task described above can be used for the bedside diagnosis of simultanagnosia. In this modification, some of the targets (e.g., As) are made to be much larger than the others (7.5–10 cm vs 2.5 cm [3–4 in. vs 1 in.] in height), and all targets are embedded among foils. Patients with simultanagnosia display a counterintuitive but characteristic tendency to miss the larger targets (Fig. 30-3B). This occurs because the information needed for the identification of the larger targets cannot be confined to the immediate line of gaze and requires the integration of visual information across multiple fixation points. The greater difficulty in the detection of the larger targets also indicates that poor acuity is not responsible for the impairment of visual function and that the problem is central rather than peripheral. The test shown in Fig. 30-3B is not by itself sufficient to diagnose simultanagnosia as some patients with a frontal network syndrome may omit the letters that appear incongruous for the size of the paper. This may happen because they lack the mental flexibility to realize that the two types of targets are symbolically identical despite being superficially different.

Bilateral parietal lesions can impair the integration of egocentric with allocentric spatial coordinates. One manifestation is *dressing apraxia*. A patient with this condition is unable to align the body axis with the axis of the garment and can be seen struggling as he or she holds a coat from its bottom or extends his or her arm into a fold of the garment rather than into its sleeve. Lesions that involve the posterior parietal cortex also lead to severe difficulties in copying simple line drawings. This is known as a *construction apraxia* and is much more severe if the lesion is in the right hemisphere. In some patients with right hemisphere lesions, the drawing difficulties are confined to the left side of the figure and represent a manifestation of hemispatial neglect; in others, there is a more universal deficit in reproducing contours and three-dimensional perspective. Impairments of route finding can be included in this group of disorders, which reflect an inability to orient the self with respect to external objects and landmarks.

Causes of Spatial Disorientation and the Posterior Cortical Atrophy Syndrome Cerebrovascular lesions and neoplasms in the right hemisphere are common causes of hemispatial neglect. Depending on the site of the lesion, a patient with neglect also may have hemiparesis, hemihypesthesia, and hemianopia on the left, but these are not invariant findings. The majority of these patients display considerable improvement of hemispatial neglect, usually within the first several weeks. Bálint’s syndrome, dressing apraxia, and route-finding impairments are more likely to result from bilateral dorsal parietal lesions; common settings for acute onset include watershed infarction between the middle and posterior cerebral artery territories, hypoglycemia, and sagittal sinus thrombosis.

A progressive form of spatial disorientation, known as the *posterior cortical atrophy* (PCA) syndrome, most commonly represents a variant of AD with unusual concentrations of neurofibrillary degeneration in the parieto-occipital cortex and the superior colliculus (Fig. 30-4). Lewy body disease (LBD), CJD, and FTLD (corticobasal degeneration type) are other possible causes. The patient displays progressive hemispatial neglect, Bálint’s syndrome, and route-finding impairments, usually accompanied by dressing and construction apraxia.

THE OCCIPITOTEMPORAL NETWORK FOR FACE AND OBJECT RECOGNITION

A patient with *prosopagnosia* cannot recognize familiar faces, including, sometimes, the reflection of their own face in the mirror. This is not a perceptual deficit because prosopagnosic patients easily can tell whether two faces are identical. Furthermore, a prosopagnosic patient who cannot recognize a familiar face by visual inspection alone can use auditory cues to reach appropriate recognition if allowed to listen to the person’s voice. The deficit in prosopagnosia is therefore modality-specific and reflects the existence of a lesion that prevents the activation of otherwise intact multimodal associative templates by relevant visual input. Prosopagnosic patients characteristically have no difficulty with the generic identification of a face as a face or a car as a car, but may not recognize the identity of an individual face or the make of an individual car. This reflects a visual recognition deficit for proprietary features that characterize individual members of an object class. When recognition problems become more generalized and extend to the generic identification of common objects, the condition is known as *visual object agnosia*. A patient with anomia cannot name the object but can describe its use. In contrast, a patient with visual agnosia is unable either to name a visually presented object or to describe its use. Face and object recognition disorders also can result from the simultanagnosia of Bálint’s syndrome, in which case they are known as *apperceptive agnosias* as opposed to the *associative agnosias* that result from inferior temporal lobe lesions.

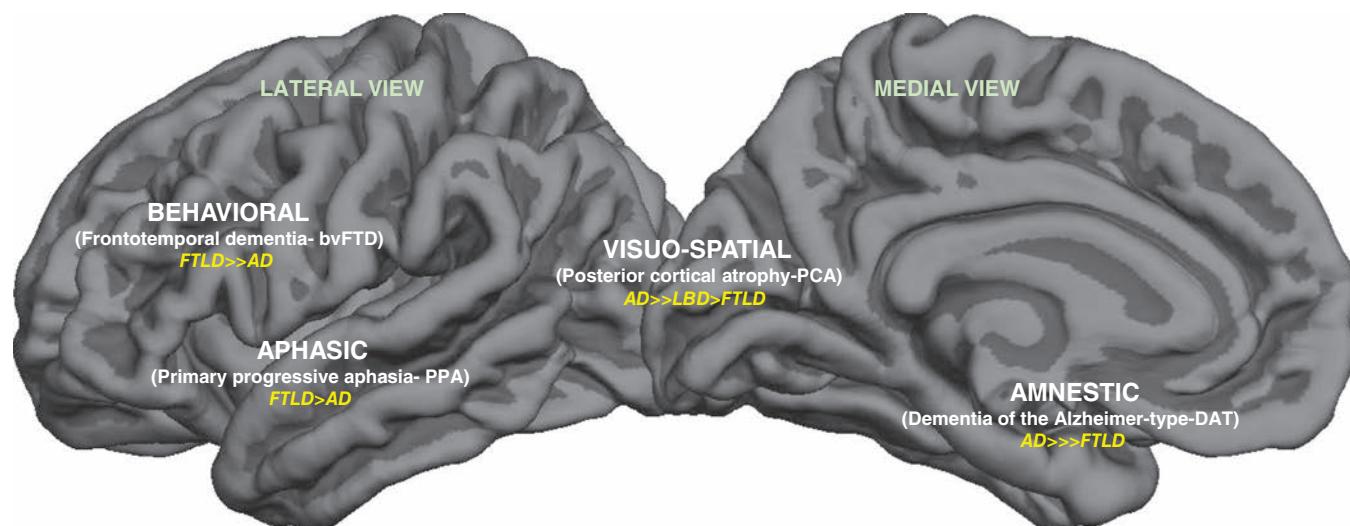


FIGURE 30-4 Four focal dementia syndromes and their most likely neuropathologic correlates. AD, Alzheimer’s disease; bvFTD, behavioral variant frontotemporal dementia; DAT, amnestic dementia of the Alzheimer type; FTLD, frontotemporal lobar degeneration (tau or TDP-43 type); LBD, Lewy body disease; PCA, posterior cortical atrophy syndrome; PPA, primary progressive aphasia.

■ CAUSES AND RELATION TO SEMANTIC DEMENTIA

The characteristic lesions in prosopagnosia and visual object agnosia of acute onset consist of bilateral infarctions in the territory of the posterior cerebral arteries that involve the fusiform gyrus. Associated deficits can include visual field defects (especially superior quadrantanopias) and a centrally based color blindness known as achromatopsia. Rarely, the responsible lesion is unilateral. In such cases, prosopagnosia is associated with lesions in the right hemisphere, and object agnosia with lesions in the left. Degenerative diseases of anterior and inferior temporal cortex can cause progressive associative prosopagnosia and object agnosia. The combination of progressive associative agnosia and a fluent aphasia with word comprehension impairment is known as *semantic dementia*. Patients with semantic dementia fail to recognize faces and objects and cannot understand the meaning of words denoting objects. This needs to be differentiated from the semantic type of PPA where there is severe impairment in understanding words that denote objects and in naming faces and objects but a relative preservation of face and object recognition. The anterior temporal lobe atrophy is usually bilateral in semantic dementia whereas it tends to affect mostly the left hemisphere in semantic PPA. Acute onset of the semantic dementia syndrome can be associated with herpes simplex encephalitis.

LIMBIC NETWORK FOR EXPLICIT MEMORY AND AMNESIA

Limbic areas (e.g., the hippocampus, amygdala, and entorhinal cortex), paralimbic areas (e.g., the cingulate gyrus, insula, temporopolar cortex, and parts of orbitofrontal regions), the anterior and medial nuclei of the thalamus, the medial and basal parts of the striatum, and the hypothalamus collectively constitute a distributed network known as the *limbic system*. The behavioral affiliations of this network can be classified into two groups. One includes the coordination of emotion, motivation, affiliative behaviors, autonomic tone, and endocrine function. These functions are under the influence of the amygdala and anterior paralimbic areas. They make up the salience network. The two neurologic conditions that most frequently interfere with this group of limbic functions are temporal lobe epilepsy and behavioral variant frontotemporal dementia (bvFTD). An additional area of specialization for the limbic network and the one that is of most relevance to clinical practice is that of declarative (explicit) memory for recent episodes and experiences. This function is under the influence of the hippocampus, entorhinal cortex, posterior paralimbic areas, and limbic nuclei of the thalamus. This part of the limbic system is also known as the Papez circuit. A disturbance of explicit memory is known as an *amnestic state*. In the absence of deficits in motivation, attention, language, or visuospatial function, the clinical diagnosis of a persistent global amnestic state is always associated with bilateral damage to the limbic network, usually within the hippocampo-entorhinal complex or the thalamus. Damage to the limbic network does not necessarily destroy memories but interferes with their conscious recall in coherent form. The individual fragments of information remain preserved despite the limbic lesions and can sustain what is known as *implicit memory*. For example, patients with amnestic states can acquire new motor or perceptual skills even though they may have no conscious knowledge of the experiences that led to the acquisition of these skills.

The memory disturbance in the amnestic state is multimodal and includes retrograde and anterograde components. The *retrograde amnesia* involves an inability to recall experiences that occurred before the onset of the amnestic state. Relatively recent events are more vulnerable to retrograde amnesia than are more remote and more extensively consolidated events. A patient who comes to the emergency room complaining that he cannot remember his or her identity but can remember the events of the previous day almost certainly does not have a neurologic cause of memory disturbance. The second and most important component of the amnestic state is the *anterograde amnesia*, which indicates an inability to store, retain, and recall new knowledge. Patients with amnestic states cannot remember what they

ate a few hours ago or the details of an important event they may have experienced in the recent past. In the acute stages, there also may be a tendency to fill in memory gaps with inaccurate, fabricated, and often implausible information. This is known as *confabulation*. Patients with the amnestic syndrome forget that they forget and tend to deny the existence of a memory problem when questioned. Confabulation is more common in cases where the underlying lesion also interferes with parts of the frontal network, as in the case of the Wernicke-Korsakoff syndrome or traumatic head injury.

■ CLINICAL EXAMINATION

A patient with an amnestic state is almost always disoriented, especially to time, and has little knowledge of current news. The anterograde component of an amnestic state can be tested with a list of four to five words read aloud by the examiner up to five times or until the patient can immediately repeat the entire list without an intervening delay. The next phase of the recall occurs after a period of 5–10 min during which the patient is engaged in other tasks. Amnestic patients fail this phase of the task and may even forget that they were given a list of words to remember. Accurate recognition of the words by multiple choice in a patient who cannot recall them indicates a less severe memory disturbance that affects mostly the retrieval stage of memory. The retrograde component of an amnesia can be assessed with questions related to autobiographical or historic events. The anterograde component of amnestic states is usually much more prominent than the retrograde component. In rare instances, occasionally associated with temporal lobe epilepsy or herpes simplex encephalitis, the retrograde component may dominate. Confusional states caused by toxic-metabolic encephalopathies and some types of frontal lobe damage lead to secondary memory impairments, especially at the stages of encoding and retrieval, even in the absence of limbic lesions. This sort of memory impairment can be differentiated from the amnestic state by the presence of additional impairments in the attention-related tasks described below in the section on the frontal lobes.

■ CAUSES, INCLUDING ALZHEIMER'S DISEASE

Neurologic diseases that give rise to an amnestic state include tumors (of the sphenoid wing, posterior corpus callosum, thalamus, or medial temporal lobe), infarctions (in the territories of the anterior or posterior cerebral arteries), head trauma, herpes simplex encephalitis, Wernicke-Korsakoff encephalopathy, autoimmune limbic encephalitis, and degenerative dementias such as AD and Pick's disease. The one common denominator of all these diseases is the presence of bilateral lesions within one or more components in the limbic network. Occasionally, unilateral left-sided hippocampal lesions can give rise to an amnestic state, but the memory disorder tends to be transient. Depending on the nature and distribution of the underlying neurologic disease, the patient also may have visual field deficits, eye movement limitations, or cerebellar findings.

The most common cause of progressive memory impairments in the elderly is AD. This is why a predominantly amnestic dementia is also known as a dementia of the Alzheimer type (DAT). A prodromal stage of DAT, when daily living activities are generally preserved, is known as amnestic mild cognitive impairment (MCI). The predilection of the entorhinal cortex and hippocampus for early neurofibrillary degeneration by typical AD pathology is responsible for the initially selective impairment of episodic memory. In time, additional impairments in language, attention, and visuospatial skills emerge as the neurofibrillary degeneration spreads to additional neocortical areas. Less frequently, amnestic dementias can also be caused by FTLD.

Transient global amnesia is a distinctive syndrome usually seen in late middle age. Patients become acutely disoriented and repeatedly ask who they are, where they are, and what they are doing. The spell is characterized by anterograde amnesia (inability to retain new information) and a retrograde amnesia for relatively recent events that occurred before the onset. The syndrome usually resolves within 24–48 h and is followed by the filling in of the period affected by the retrograde amnesia, although there is persistent loss of memory for the events that occurred during the ictus. Recurrences are noted in

~20% of patients. Migraine, temporal lobe seizures, and perfusion abnormalities in the posterior cerebral territory have been postulated as causes of transient global amnesia. The absence of associated neurologic findings occasionally may lead to the incorrect diagnosis of a psychiatric disorder.

THE PREFRONTAL NETWORK FOR EXECUTIVE FUNCTION AND BEHAVIOR

The frontal lobes can be subdivided into motor-premotor, dorsolateral prefrontal, medial prefrontal, and orbitofrontal components. The terms *frontal lobe syndrome* and *prefrontal cortex* refer only to the last three of these four components. These are the parts of the cerebral cortex that show the greatest phylogenetic expansion in primates, especially in humans. The dorsolateral prefrontal, medial prefrontal, and orbitofrontal areas, along with the subcortical structures with which they are interconnected (i.e., the head of the caudate and the dorsomedial nucleus of the thalamus), collectively make up a large-scale network that coordinates exceedingly complex aspects of human cognition and behavior. The prefrontal network overlaps with the salience network through the anterior cingulate gyrus and parts of the orbitofrontal region. Impairments of social conduct and empathy seen in neurodegenerative frontal dementias (such as bvFTD) are attributed to pathology of the prefrontal and salience networks.

The prefrontal network plays an important role in behaviors that require multitasking and the integration of thought with emotion. Cognitive operations impaired by prefrontal cortex lesions often are referred to as “executive functions.” The most common clinical manifestations of damage to the prefrontal network take the form of two relatively distinct syndromes. In the *frontal abulia syndrome*, the patient shows a loss of initiative, creativity, and curiosity and displays a pervasive emotional blandness, apathy, and lack of empathy. In the *frontal disinhibition syndrome*, the patient becomes socially disinhibited and shows severe impairments of judgment, insight, foresight, and the ability to mind rules of conduct. The dissociation between intact intellectual function and a total lack of even rudimentary common sense is striking. Despite the preservation of all essential memory functions, the patient cannot learn from experience and continues to display inappropriate behaviors without appearing to feel emotional pain, guilt, or regret when those behaviors repeatedly lead to disastrous consequences. The impairments may emerge only in real-life situations when behavior is under minimal external control and may not be apparent within the structured environment of the medical office. Testing judgment by asking patients what they would do if they detected a fire in a theater or found a stamped and addressed envelope on the road is not very informative because patients who answer these questions wisely in the office may still act very foolishly in real-life settings. The physician must therefore be prepared to make a diagnosis of frontal lobe disease based on historic information alone even when the mental state is quite intact in the office examination.

CLINICAL EXAMINATION

The emergence of developmentally primitive reflexes, also known as frontal release signs, such as grasping (elicited by stroking the palm) and sucking (elicited by stroking the lips) are seen primarily in patients with large structural lesions that extend into the premotor components of the frontal lobes or in the context of metabolic encephalopathies. The vast majority of patients with prefrontal lesions and frontal lobe behavioral syndromes do not display these reflexes. Damage to the frontal lobe disrupts a variety of attention-related functions, including working memory (the transient online holding and manipulation of information), concentration span, the effortful scanning and retrieval of stored information, the inhibition of immediate but inappropriate responses, and mental flexibility. Digit span (which should be seven forward and five reverse) is decreased, reflecting poor working memory; the recitation of the months of the year in reverse order (which should take <15 s) is slowed as another indication of poor working memory; and the fluency in producing words starting with the letter a, f, or s that can be generated in 1 min (normally ≥12 per letter) is diminished even in nonaphasic patients, indicating an impairment in

the ability to search and retrieve information from long-term stores. In “go-no go” tasks (where the instruction is to raise the finger upon hearing one tap but keep it still upon hearing two taps), the patient shows a characteristic inability to inhibit the response to the “no go” stimulus. Mental flexibility (tested by the ability to shift from one criterion to another in sorting or matching tasks) is impoverished; distractibility by irrelevant stimuli is increased; and there is a pronounced tendency for impersistence and perseveration. The ability for abstracting similarities and interpreting proverbs is also undermined.

The attentional deficits disrupt the orderly registration and retrieval of new information and lead to *secondary* deficits of explicit memory. The distinction of the underlying neural mechanisms is illustrated by the observation that severely amnestic patients who cannot remember events that occurred a few minutes ago may have intact if not superior working memory capacity as shown in tests of digit span. The use of the term *memory* to designate two completely different mental faculties is confusing. Working memory depends on the on-line holding of information for brief periods of time, whereas explicit memory depends on the off-line storage and subsequent retrieval of the information.

CAUSES: TRAUMA, NEOPLASM, AND FRONTOTEMPORAL DEMENTIA

The abulic syndrome tends to be associated with damage in dorsolateral or dorsomedial prefrontal cortex, and the disinhibition syndrome with damage in orbitofrontal or ventromedial cortex. These syndromes tend to arise almost exclusively after bilateral lesions. Unilateral lesions confined to the prefrontal cortex may remain silent until the pathology spreads to the other side; this explains why thromboembolic CVA is an unusual cause of the frontal lobe syndrome. When behavioral syndromes of the frontal network arise in conjunction with asymmetric disease, the lesion tends to be predominantly on the right side of the brain. Common settings for frontal lobe syndromes include head trauma, ruptured aneurysms, hydrocephalus, tumors (including metastases, glioblastoma, and falx or olfactory groove meningiomas), and focal degenerative diseases, especially FTLD. The most prominent neurodegenerative frontal syndrome is bvFTD. In many patients with bvFTD, the atrophy includes orbitofrontal cortex and also extends into the anterior temporal lobes, insula, and anterior cingulate cortex. Occasionally, atrophy predominantly in the right anterior temporal lobe presents with the bvFTD syndrome. The behavioral changes in these patients can range from apathy to shoplifting, compulsive gambling, sexual indiscretions, remarkable lack of common sense, new ritualistic behaviors, and alterations in dietary preferences, usually leading to increased taste for sweets or rigid attachment to specific food items. In many patients with AD, neurofibrillary degeneration eventually spreads to prefrontal cortex and gives rise to components of the frontal lobe syndrome, but almost always on a background of severe memory impairment. Rarely, the bvFTD syndrome can arise in isolation in the context of an atypical form of AD pathology.

Lesions in the caudate nucleus or in the dorsomedial nucleus of the thalamus (subcortical components of the prefrontal network) also can produce a frontal lobe syndrome affecting mostly executive functions. This is one reason why the changes in mental state associated with degenerative basal ganglia diseases such as Parkinson’s disease and Huntington’s disease display components of the frontal lobe syndrome. Bilateral multifocal lesions of the cerebral hemispheres, none of which are individually large enough to cause specific cognitive deficits such as aphasia and neglect, can collectively interfere with the connectivity and therefore integrating (executive) function of the prefrontal cortex. A frontal lobe syndrome, usually of the abulic form, is therefore the single most common behavioral profile associated with a variety of bilateral multifocal brain diseases, including metabolic encephalopathy, multiple sclerosis, and vitamin B₁₂ deficiency, among others. Many patients with the clinical diagnosis of a frontal lobe syndrome tend to have lesions that do not involve prefrontal cortex but involve either the subcortical components of the prefrontal network or its connections with other parts of the brain. To avoid making a diagnosis of “frontal lobe syndrome” in a patient with no evidence of frontal cortex disease, it is advisable to use the diagnostic term *frontal network syndrome*, with the

understanding that the responsible lesions can lie anywhere within this distributed network. A patient with frontal lobe disease raises potential dilemmas in differential diagnosis: the abulia and blandness may be misinterpreted as depression, and the disinhibition as idiopathic mania or acting out. Appropriate intervention may be delayed while a treatable tumor keeps expanding.

CARING FOR PATIENTS WITH DEFICITS OF HIGHER CEREBRAL FUNCTION

Spontaneous improvement of cognitive deficits following stroke or trauma is common. It is most rapid in the first few weeks but may continue for up to 2 years, especially in young individuals with single brain lesions. Some of the initial deficits in such cases appear to arise from remote dysfunction (diaschisis) in brain regions that are interconnected with the site of initial injury. Improvement in these patients may reflect, at least in part, a normalization of the remote dysfunction. Other mechanisms may involve functional reorganization in surviving neurons adjacent to the injury or the compensatory use of homologous structures, e.g., the right superior temporal gyrus with recovery from Wernicke's aphasia. In contrast, neurodegenerative diseases show a progression of impairment but at rates that vary greatly from patient to patient.

Pharmacologic and Nonpharmacologic Interventions Some of the deficits described in this chapter are so complex that they may bewilder not only the patient and family but also the physician. The care of patients with such deficits requires a careful evaluation of the history, cognitive test results, and diagnostic procedures. Each piece of information needs to be interpreted cautiously and placed in context. A complaint of "poor memory," for example, may reflect an anomia; poor scores on a learning task may reflect a weakness of attention rather than explicit memory; a report of depression or indifference may reflect impaired prosody rather than a change in mood or empathy; jocularity may arise from poor insight rather than good mood. Although there are few well-controlled studies, several nonpharmacologic interventions have been used to treat higher cortical deficits. These include speech therapy for aphasias, behavioral modification for compartmental disorders, and cognitive training for visuospatial disorientation and amnestic syndromes. More practical interventions, usually delivered through occupational therapy, aim to improve daily living activities through assistive devices and modifications of the home environment. Determining driving competence is challenging, especially in the early stages of dementing diseases. An on-the-road driving test and reports from family members may help time decisions related to this very important activity. In neurodegenerative conditions such as PPA, transcranial magnetic (or direct current) stimulation has had mixed success in eliciting symptomatic improvement. The goal is to activate remaining neurons at sites of atrophy or in unaffected regions of the contralateral hemisphere. Depression and sleep disorders can intensify the cognitive disorders and should be treated with appropriate modalities. If neuroleptics become absolutely necessary for the control of agitation, atypical neuroleptics are preferable because of their lower extrapyramidal side effects. Treatment with neuroleptics in elderly patients with dementia requires weighing the potential benefits against the potentially serious side effects. This is especially relevant to the case of patients with Lewy body dementia, who can be unusually sensitive to side effects.

As in all other branches of medicine, a crucial step in patient care is to identify the underlying cause of the impairment. This is easily done in cases of CVA, head trauma, or encephalitis but becomes particularly challenging in the dementias because the same progressive clinical syndrome can be caused by one of several neuropathologic entities. The advent of imaging, blood, and cerebrospinal fluid biomarkers now makes it possible to address this question with reasonable success and to make specific diagnoses of AD, LBD, CJD, and FTLD. A specific etiologic diagnosis allows the physician to recommend medications or clinical trials that are the most appropriate for the underlying disease process. A clinical assessment that identifies the principal domain of behavioral and cognitive impairment followed by the judicious use of

biomarker information to surmise the nature of the underlying disease allows a personalized approach to patients with higher cognitive impairment.

FURTHER READING

- GHETTI B et al: *Frontotemporal Dementias: Emerging Milestones of the 21st Century*. New York, Springer, 2021.
 HENRY ML et al: Retraining speech production and fluency in non-fluent/agrammatic primary progressive aphasia. *Brain* 141:1799, 2018.
 MESULAM M-M: Behavioral neuroanatomy: Large-scale networks, association cortex, frontal syndromes, the limbic system and hemispheric specialization, in *Principles of Behavioral and Cognitive Neurology*, M-M Mesulam (ed). New York, Oxford University Press, 2000, pp 1-120.
 MESULAM M-M et al: Word comprehension in temporal cortex and Wernicke area: A PPA perspective. *Neurology* 92:e224, 2019.
 MILLER BL, BOEVE BF (eds): *The Behavioral Neurology of Dementia*, 2nd ed. Cambridge, Cambridge University Press, 2017.

31

Sleep Disorders

Thomas E. Scammell, Clifford B. Saper,
 Charles A. Czeisler



Disturbed sleep is one of the most common health complaints that physicians encounter. More than one-half of adults in the United States experience at least intermittent sleep disturbance, and only 30% of adult Americans report consistently obtaining a sufficient amount of sleep. The National Academy of Medicine has estimated that 50–70 million Americans suffer from a chronic disorder of sleep and wakefulness, which can adversely affect daytime functioning as well as physical and mental health. A high prevalence of sleep disorders across all cultures is also now increasingly recognized, and these problems are expected to further increase in the years ahead as the global population ages. Over the last 30 years, the field of sleep medicine has emerged as a distinct specialty in response to the impact of sleep disorders and sleep deficiency on overall health. Nonetheless, over 80% of patients with sleep disorders remain undiagnosed and untreated—costing the U.S. economy over \$400 billion annually in increased health care costs, lost productivity, accidents and injuries, and leading to the development of workplace-based sleep health education and sleep disorders screening programs designed to address this unmet medical need.

PHYSIOLOGY OF SLEEP AND WAKEFULNESS

Most adults need 7–9 h of sleep per night to promote optimal health, although the timing, duration, and internal structure of sleep vary among individuals. In the United States, adults tend to have one consolidated sleep episode each night, although in some cultures sleep may be divided into a mid-afternoon nap and a shortened night sleep. This pattern changes considerably over the life span, as infants and young children sleep considerably more than older people, while individuals >70 years of age sleep on average about an hour less than young adults.

The stages of human sleep are defined on the basis of characteristic patterns in the electroencephalogram (EEG), the electrooculogram (EOG—a measure of eye-movement activity), and the surface electromyogram (EMG) measured on the chin, neck, and legs. The continuous recording of these electrophysiologic parameters to define sleep and wakefulness is termed *polysomnography*.

Polysomnographic profiles define two basic states of sleep: (1) rapid eye movement (REM) sleep and (2) non-rapid eye movement (NREM)

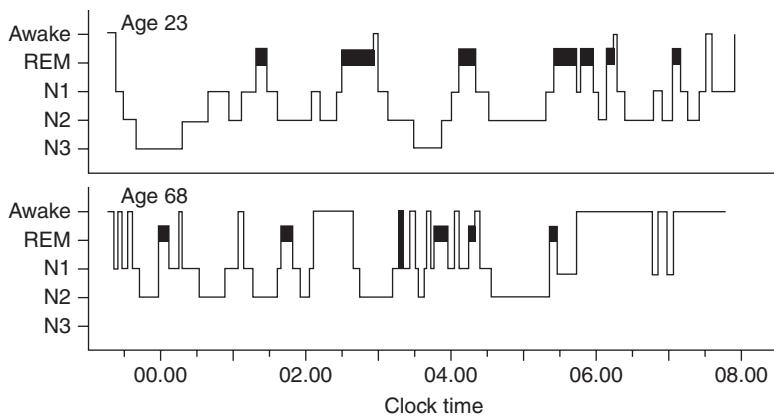


FIGURE 31-1 Wake-sleep architecture. Alternating stages of wakefulness, the three stages of non-rapid eye movement sleep (N1–N3), and rapid eye movement (REM) sleep (solid bars) occur over the course of the night for representative young and older adult men. Characteristic features of sleep in older people include reduction of N3 slow-wave sleep, frequent spontaneous awakenings, early sleep onset, and early morning awakening.

sleep. NREM sleep is further subdivided into three stages: N1, N2, and N3, characterized by an increasing threshold for arousal and slowing of the cortical EEG. REM sleep is characterized by a low-amplitude, mixed-frequency EEG, similar to NREM stage N1 sleep, and an EOG pattern of REMs that tend to occur in flurries or bursts. EMG activity is absent in nearly all skeletal muscles except those involved in respiration, reflecting the brainstem-mediated muscle paralysis that is characteristic of REM sleep.

ORGANIZATION OF HUMAN SLEEP

Normal nocturnal sleep in adults displays a consistent organization from night to night (Fig. 31-1). After sleep onset, sleep usually progresses through NREM stages N1–N3 sleep within 45–60 min. NREM stage N3 sleep (also known as slow-wave sleep) predominates in the first third of the night and comprises 15–25% of total nocturnal sleep time in young adults. Sleep deprivation increases the rapidity of sleep onset and both the intensity and amount of slow-wave sleep.

The first REM sleep episode usually occurs in the second hour of sleep. NREM and REM sleep alternate through the night with an average period of 90–110 min (the “ultradian” sleep cycle). Overall, in a healthy young adult, REM sleep constitutes 20–25% of total sleep, and NREM stages N1 and N2 constitute 50–60%.

Age has a profound impact on sleep state organization (Fig. 31-1). N3 sleep is most intense and prominent during childhood, decreasing with puberty and across the second and third decades of life. In older adults, N3 sleep may be completely absent, and the remaining NREM sleep typically becomes more fragmented, with frequent awakenings from NREM sleep. It is the increased frequency of awakenings, rather than a decreased ability to fall back asleep, that accounts for the increased wakefulness during the sleep episode in older people. While REM sleep may account for 50% of total sleep time in infancy, the percentage falls off sharply over the first postnatal year as a mature REM-NREM cycle develops; thereafter, REM sleep occupies about 25% of total sleep time.

Sleep deprivation degrades cognitive performance, particularly on tests that require continual vigilance. Paradoxically, older people are less vulnerable than young adults to the neurobehavioral performance impairment induced by acute sleep deprivation, maintaining their reaction time and sustaining vigilance with fewer lapses of attention. However, it is more difficult for older adults to obtain recovery sleep after staying awake all night, as the ability to sleep during the daytime declines with age.

After sleep deprivation, NREM sleep generally recovers first, followed by REM sleep. However, because REM sleep tends to be most prominent in the second half of the night, sleep truncation (e.g., by an alarm clock) results in selective REM sleep deprivation. This may increase REM sleep pressure to the point where the first REM sleep may occur much earlier in the nightly sleep episode. Because several

disorders (see below) also cause sleep fragmentation, it is important that the patient have sufficient sleep opportunity (at least 8 h per night) for several nights prior to a diagnostic polysomnogram.

There is growing evidence that inadequate sleep in humans is associated with glucose intolerance that may contribute to the development of diabetes, obesity, and the metabolic syndrome, as well as impaired immune responses, accelerated atherosclerosis, and increased risk of cardiac disease, cognitive impairment, Alzheimer’s disease, and stroke. For these reasons, the National Academy of Medicine declared sleep deficiency and sleep disorders “an unmet public health problem.”

WAKE AND SLEEP ARE REGULATED BY BRAIN CIRCUITS

Two principal neural systems govern the expression of sleep and wakefulness. The ascending arousal system, illustrated in green in Fig. 31-2, consists of clusters of nerve cells extending from the upper pons to the hypothalamus and basal forebrain that activate the cerebral cortex, thalamus

(which is necessary to relay sensory information to the cortex), and other forebrain regions. The ascending arousal neurons use monoamines (norepinephrine, dopamine, serotonin, and histamine), glutamate, or acetylcholine as neurotransmitters to activate their target neurons. Some basal forebrain neurons use γ -aminobutyric acid (GABA) to inhibit cortical inhibitory interneurons, thus promoting arousal. Additional wake-promoting neurons in the hypothalamus use the peptide neurotransmitter orexin (also known as hypocretin, shown in Fig. 31-2 in blue) to reinforce activity in the other arousal cell groups.

Damage to the arousal system at the level of the rostral pons and lower midbrain causes coma, indicating that the ascending arousal influence from this level is critical in maintaining wakefulness. Injury to the hypothalamic branch of the arousal system causes profound sleepiness but usually not coma. Specific loss of the orexin neurons produces the sleep disorder narcolepsy (see below). Isolated damage to the thalamus causes loss of the content of wakefulness, known as a persistent vegetative state, but wake-sleep cycles are largely preserved.

The arousal system is turned off during sleep by inhibitory inputs from cell groups in the sleep-promoting system, shown in Fig. 31-2 in red. These neurons in the preoptic area and pons use GABA to inhibit the arousal system. Additional neurons in the lateral hypothalamus containing the peptide melanin-concentrating hormone promote REM sleep. Many sleep-promoting neurons are themselves inhibited by inputs from the arousal system. This mutual inhibition between the arousal- and sleep-promoting systems forms a neural circuit akin to what electrical engineers call a “flip-flop switch.” A switch of this type tends to promote rapid transitions between the on (wake) and off (sleep) states, while avoiding intermediate states. The relatively rapid transitions between waking and sleeping states, as seen in the EEG of humans and animals, is consistent with this model.

Neurons in the ventrolateral preoptic nucleus, one of the key sleep-promoting sites, are lost during normal human aging, correlating with reduced ability to maintain sleep (sleep fragmentation). The ventrolateral preoptic neurons are also injured in Alzheimer’s disease, which may in part account for the poor sleep quality in those patients.

Transitions between NREM and REM sleep appear to be governed by a similar switch in the brainstem. GABAergic REM-Off neurons have been identified in the lower midbrain that inhibit REM-On neurons in the upper pons. The REM-On group contains both GABAergic neurons that inhibit the REM-Off group (thus satisfying the conditions for a REM sleep flip-flop switch) as well as glutamatergic neurons that project widely in the central nervous system (CNS) to cause the key phenomena associated with REM sleep. REM-On neurons that project to the medulla and spinal cord activate inhibitory (GABA and glycine-containing) interneurons, which in turn hyperpolarize the motor neurons, producing the paralysis of REM sleep. REM-On neurons that project to the forebrain may be important in producing dreams.

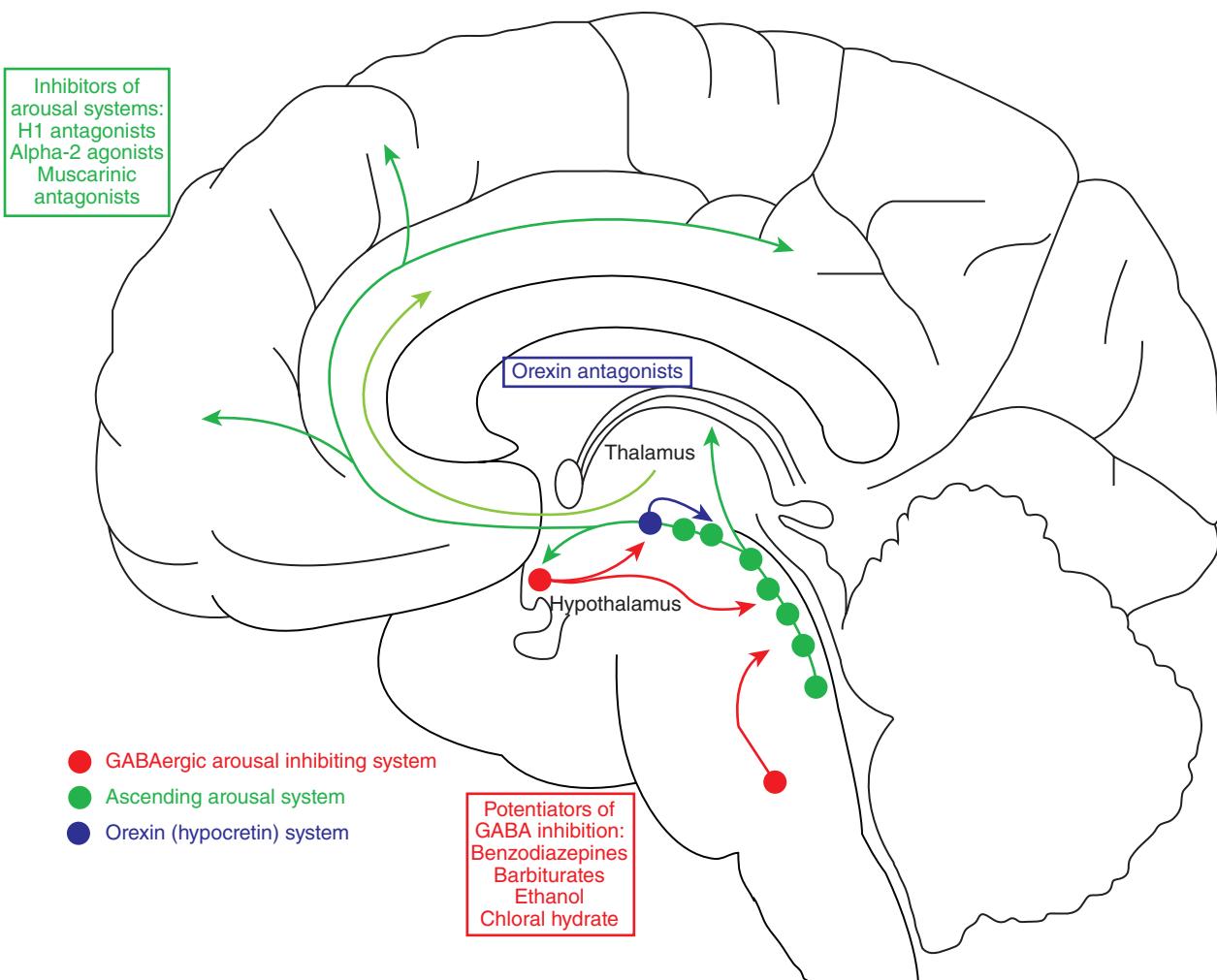


FIGURE 31-2 Relationship of drugs for insomnia with wake-sleep systems. The arousal system in the brain (green) includes monoaminergic, glutamatergic, and cholinergic neurons in the brainstem that activate neurons in the hypothalamus, thalamus, basal forebrain, and cerebral cortex. Orexin neurons (blue) in the hypothalamus, which are lost in narcolepsy, reinforce and stabilize arousal by activating other components of the arousal system. The sleep-promoting system (red) consists of GABAergic neurons in the preoptic area and brainstem that inhibit the components of the arousal system, thus allowing sleep to occur. Drugs used to treat insomnia include those that block the effects of arousal system neurotransmitters (green and blue) and those that enhance the effects of γ -aminobutyric acid (GABA) produced by the sleep system (red).

The REM sleep switch receives cholinergic input, which favors transitions to REM sleep, and monoaminergic (norepinephrine and serotonin) input that prevents REM sleep. As a result, drugs that increase monoamine tone (e.g., serotonin or norepinephrine reuptake inhibitors) tend to reduce the amount of REM sleep. Damage to the neurons that promote REM sleep paralysis can produce REM sleep behavior disorder, a condition in which patients act out their dreams (see below).

SLEEP-WAKE CYCLES ARE DRIVEN BY HOMEOSTATIC, ALLOSTATIC, AND CIRCADIAN INPUTS

The gradual increase in sleep drive with prolonged wakefulness, followed by deeper slow-wave sleep and prolonged sleep episodes, demonstrates that there is a *homeostatic* mechanism that regulates sleep. The neurochemistry of sleep homeostasis is only partially understood, but with prolonged wakefulness, adenosine levels rise in parts of the brain. Adenosine may act through A1 receptors to directly inhibit many arousal-promoting brain regions. In addition, adenosine promotes sleep through A2a receptors; blockade of these receptors by caffeine is one of the chief ways in which people fight sleepiness. Other humoral factors, such as prostaglandin D₂, have also been implicated in this process. Both adenosine and prostaglandin D₂ activate the sleep-promoting neurons in the ventrolateral preoptic nucleus.

Allostasis is the physiologic response to a challenge such as physical danger or psychological threat that cannot be managed by homeostatic mechanisms. These stress responses can severely impact the need for

and ability to sleep. For example, insomnia is very common in patients with anxiety and other psychiatric disorders. Stress-induced insomnia is even more common, affecting most people at some time in their lives. Positron emission tomography (PET) studies in patients with chronic insomnia show hyperactivation of components of the ascending arousal system, as well as their limbic system targets in the forebrain (e.g., cingulate cortex and amygdala). The limbic areas are not only targets for the arousal system, but they also send excitatory outputs back to the arousal system, which contributes to a vicious cycle of anxiety about insomnia that makes it more difficult to sleep. Approaches to treating insomnia may employ drugs that either inhibit the output of the ascending arousal system (green and blue in Fig. 31-2) or potentiate the output of the sleep-promoting system (red in Fig. 31-2). However, behavioral approaches (cognitive behavioral therapy [CBT] and sleep hygiene) that may reduce forebrain limbic activity at bedtime are often the best long-term treatment.

Sleep is also regulated by a strong *circadian* timing signal, driven by the suprachiasmatic nuclei (SCN) of the hypothalamus, as described below. The SCN sends outputs to key sites in the hypothalamus, which impose 24-h rhythms on a wide range of behaviors and body systems, including the wake-sleep cycle.

PHYSIOLOGY OF CIRCADIAN RHYTHMICITY

The wake-sleep cycle is the most evident of many 24-h rhythms in humans. Prominent daily variations also occur in endocrine, thermoregulatory, cardiac, pulmonary, renal, immune, gastrointestinal, and neurobehavioral functions. In evaluating daily rhythms in humans,

it is important to distinguish between diurnal components passively evoked by periodic environmental or behavioral changes (e.g., the increase in blood pressure and heart rate that occurs upon assumption of the upright posture) and circadian rhythms actively driven by an endogenous oscillatory process (e.g., the circadian variations in adrenal cortisol and pineal melatonin secretion that persist across a variety of environmental and behavioral conditions).

At the cellular level, endogenous circadian rhythmicity is driven by self-sustaining feedback loops. While it is now recognized that most cells in the body have circadian clocks that regulate diverse physiologic processes, these clocks in different tissues, or even in different cells in the same tissue, when placed in isolation in a tissue explant are unable to maintain the long-term synchronization with each other that is required to produce useful 24-h rhythms aligned with the external light-dark cycle. The only tissue that maintains this rhythm *in vitro* is the SCN, whose neurons are interconnected with one another in such a way as to produce a near-24-h synchronous rhythm of neural activity even in prolonged slice culture. SCN neurons are located just above the optic chiasm in the hypothalamus, from which they receive visual input to synchronize them with the external world, and they have outputs to transmit that signal to the rest of the body. Bilateral destruction of the SCN results in a loss of most endogenous circadian rhythms including wake-sleep behavior and rhythms in endocrine and metabolic systems. The genetically determined period of this endogenous neural oscillator, which averages ~24.15 h in humans, is normally synchronized to the 24-h period of the environmental light-dark cycle through direct input from intrinsically photosensitive ganglion cells in the retina to the SCN. Humans are exquisitely sensitive to the resetting effects of light, particularly the shorter wavelengths (~460–500 nm) in the blue part of the visible spectrum. Small differences in circadian period contribute to variations in diurnal preference. Changes in homeostatic sleep regulation may underlie age-related changes in sleep-wake timing.

The timing and internal architecture of sleep are directly coupled to the output of the endogenous circadian pacemaker. Paradoxically, the endogenous circadian rhythm for wake propensity peaks just before the habitual bedtime, whereas that of sleep propensity peaks near the habitual wake time. These rhythms are thus timed to oppose the rise of sleep tendency throughout the usual waking day and the decline of sleep propensity during the habitual sleep episode, respectively, thus promoting consolidated sleep and wakefulness. Misalignment of the endogenous circadian pacemaker with the desired wake-sleep cycle can, therefore, induce insomnia, decrease alertness, and impair performance, posing health problems for night-shift workers and airline travelers.

■ BEHAVIORAL AND PHYSIOLOGIC CORRELATES OF SLEEP STATES AND STAGES

Polysomnographic staging of sleep correlates with behavioral changes during specific states and stages. During the transitional state (stage N1) between wakefulness and deeper sleep, individuals may respond to faint auditory or visual signals. Formation of short-term memories is inhibited at the onset of NREM stage N1 sleep, which may explain why individuals aroused from that transitional sleep stage frequently lack situational awareness. After sleep deprivation, such transitions may intrude upon behavioral wakefulness notwithstanding attempts to remain continuously awake (for example, see “Shift-Work Disorder,” below).

Subjects awakened from REM sleep recall vivid dream imagery >80% of the time, especially later in the night. Less vivid imagery may also be reported after NREM sleep interruptions. Certain disorders may occur during specific sleep stages and are described below under “Parasomnias.” These include sleepwalking, night terrors, and enuresis (bed wetting), which occur most commonly in children during deep (N3) NREM sleep, and REM sleep behavior disorder, which occurs mainly among older men who fail to maintain full paralysis during REM sleep, and often call out, thrash around, or even act out fragments of dreams.

All major physiologic systems are influenced by sleep. Blood pressure and heart rate decrease during NREM sleep, particularly during N3 sleep. During REM sleep, bursts of eye movements are associated

with large variations in both blood pressure and heart rate mediated by the autonomic nervous system. Cardiac dysrhythmias may occur selectively during REM sleep. Respiratory function also changes. In comparison to relaxed wakefulness, respiratory rate becomes slower but more regular during NREM sleep (especially N3 sleep) and becomes irregular during bursts of eye movements in REM sleep. Decreases in minute ventilation during NREM sleep are out of proportion to the decrease in metabolic rate, resulting in a slightly higher PCO_2 .

Within the brain itself, neurotransmission is supported by ion gradients across the cell membranes of neurons and astrocytes. These ion flows are accompanied by increases in intracellular volume, so that during wake, there is very little extracellular space in the brain. During sleep, intracellular volume is reduced, resulting in increased extracellular space, which has higher calcium and lower potassium concentrations, supporting hyperpolarization and reduced firing of neurons. This expansion of the extracellular space during sleep increases diffusion of substances that accumulate extracellularly, like β -amyloid peptide, enhancing their clearance from the brain via cerebrospinal fluid (CSF) flow. Recent evidence suggests that lack of adequate sleep may contribute to extracellular accumulation of β -amyloid peptide, a key step in the pathogenesis of Alzheimer’s disease.

Endocrine function also varies with sleep. N3 sleep is associated with secretion of growth hormone in men, while sleep in general is associated with augmented secretion of prolactin in both men and women. Sleep has a complex effect on the secretion of luteinizing hormone (LH): during puberty, sleep is associated with increased LH secretion, whereas sleep in postpubertal women inhibits LH secretion in the early follicular phase of the menstrual cycle. Sleep onset (and probably N3 sleep) is associated with inhibition of thyroid-stimulating hormone and of the adrenocorticotrophic hormone-cortisol axis, an effect that is superimposed on the prominent circadian rhythms in the two systems.

The pineal hormone melatonin is secreted predominantly at night in both day- and night-active species, reflecting the direct modulation of pineal activity by the SCN via the sympathetic nervous system, which innervates the pineal gland. Melatonin secretion does not require sleep, but melatonin secretion is inhibited by ambient light, an effect mediated by the neural connection from the retina to the pineal gland via the SCN. In humans, sleep efficiency is highest when sleep coincides with endogenous melatonin secretion. When endogenous melatonin levels are low, such as during the biological day or at the desired bedtime in people with delayed sleep-wake phase disorder (DSWPD), administration of exogenous melatonin can hasten sleep onset and increase sleep efficiency, but it does not increase sleep efficiency if administered when endogenous melatonin levels are elevated. This may explain why melatonin is often ineffective in the treatment of patients with primary insomnia. On the other hand, patients with sympathetic denervation of the pineal gland, such as occurs in cervical spinal cord injury or in patients with Parkinson’s disease, often have low melatonin levels, and administration of melatonin (3 mg 30 min before bedtime) may help them sleep.

Sleep is accompanied by alterations of thermoregulatory function. NREM sleep is associated with an increase in the firing of warm-responsive neurons in the preoptic area and a fall in body temperature; conversely, skin warming without increasing core body temperature has been found to increase NREM sleep. REM sleep is associated with reduced thermoregulatory responsiveness.

DISORDERS OF SLEEP AND WAKEFULNESS

APPROACH TO THE PATIENT

Sleep Disorders

Patients may seek help from a physician because of: (1) sleepiness or tiredness during the day; (2) difficulty initiating or maintaining sleep at night (insomnia); or (3) unusual behaviors during sleep itself (parasomnias).

Obtaining a careful history is essential. In particular, the duration, severity, and consistency of the symptoms are important, along with the patient's estimate of the consequences of the sleep disorder on waking function. Information from a bed partner or family member is often helpful because some patients may be unaware of symptoms such as heavy snoring or may underreport symptoms such as falling asleep at work or while driving. Physicians should inquire about when the patient typically goes to bed, when they fall asleep and wake up, whether they awaken during sleep, whether they feel rested in the morning, and whether they nap during the day. Depending on the primary complaint, it may be useful to ask about snoring, witnessed apneas, restless sensations in the legs, movements during sleep, depression, anxiety, and behaviors around the sleep episode. The physical examination may provide evidence of a small airway, large tonsils, or a neurologic or medical disorder that contributes to the main complaint.

It is important to remember that, rarely, seizures may occur exclusively during sleep, mimicking a primary sleep disorder; such sleep-related seizures typically occur during episodes of NREM sleep and may take the form of generalized tonic-clonic movements (sometimes with urinary incontinence or tongue biting) or stereotyped movements in partial complex epilepsy ([Chap. 418](#)).

It is often helpful for the patient to complete a daily sleep log for 1–2 weeks to define the timing and amounts of sleep. When relevant, the log can also include information on levels of alertness, work times, and drug and alcohol use, including caffeine and hypnotics.

Polysomnography is necessary for the diagnosis of several disorders such as sleep apnea, narcolepsy, and periodic limb movement disorder (PLMD). A conventional polysomnogram performed in a clinical sleep laboratory allows measurement of sleep stages, respiratory effort and airflow, oxygen saturation, limb movements, heart rhythm, and additional parameters. A home sleep test usually focuses on just respiratory measures and is helpful in patients with a moderate to high likelihood of having obstructive sleep apnea. The multiple sleep latency test (MSLT) is used to measure a patient's propensity to sleep during the day and can provide crucial evidence for diagnosing narcolepsy and some other causes of sleepiness. The maintenance of wakefulness test is used to measure a patient's ability to sustain wakefulness during the daytime and can provide important evidence for evaluating the efficacy of therapies for improving sleepiness in conditions such as narcolepsy and obstructive sleep apnea.

EVALUATION OF DAYTIME SLEEPINESS

Up to 25% of the adult population has persistent daytime sleepiness that impairs an individual's ability to perform optimally in school, at work, while driving, and in other conditions that require alertness. Sleepy students often have trouble staying alert and performing well in school, and sleepy adults struggle to stay awake and focused on their work. More than half of Americans have fallen asleep while driving. An estimated 1.2 million motor vehicle crashes per year are due to drowsy drivers, causing about 20% of all serious crash injuries and deaths. One need not fall asleep to have a motor vehicle crash, as the inattention and slowed responses of drowsy drivers are major contributors. Twenty-four hours of continuous wakefulness impairs reaction time as much as a blood alcohol concentration of 0.10 g/dL (which is legally drunk in all 50 states).

Identifying and quantifying sleepiness can be challenging. First, patients may describe themselves as "sleepy," "fatigued," or "tired," and the meanings of these words may differ between patients. For clinical purposes, it is best to use the term "sleepiness" to describe a propensity to fall asleep, whereas "fatigue" is best used to describe a feeling of low physical or mental energy but without a tendency to actually sleep. Sleepiness is usually most evident when the patient is sedentary, whereas fatigue may interfere with more active pursuits. Sleepiness generally occurs with disorders that reduce the quality or quantity of sleep or that interfere with the neural mechanisms of arousal, whereas fatigue is more common in inflammatory disorders such as cancer, multiple sclerosis ([Chap. 444](#)), fibromyalgia ([Chap. 373](#)), chronic fatigue syndrome ([Chap. 450](#)), or endocrine deficiencies such as hypothyroidism ([Chap. 383](#)) or Addison's disease ([Chap. 386](#)). Second, sleepiness can affect judgment in a manner analogous to ethanol, such that patients may have limited insight into the condition and the extent of their functional impairment. Finally, patients may be reluctant to admit that sleepiness is a problem because they may have become unfamiliar with feeling fully alert, and because sleepiness is sometimes viewed pejoratively as reflecting poor motivation or bad sleep habits.

Table 31-1 outlines the diagnostic and therapeutic approach to the patient with a complaint of excessive daytime sleepiness.

To determine the extent and impact of sleepiness on daytime function, it is helpful to ask patients about the occurrence of sleep episodes during normal waking hours, both intentional and unintentional. Specific areas to be addressed include the occurrence of inadvertent sleep episodes while driving or in other safety-related settings, sleepiness while at work or school (and its impact on performance), and the effect of sleepiness on social and family life. Standardized questionnaires such as the Epworth Sleepiness Scale are often used clinically to measure sleepiness.

TABLE 31-1 Evaluation of the Patient with Excessive Daytime Sleepiness

Findings on History and Physical Examination	Diagnostic Evaluation	Diagnosis	Therapy
Difficulty waking in the morning, rebound sleep on weekends and vacations with improvement in sleepiness	Sleep log	Insufficient sleep	Sleep education and behavioral modification to increase amount of sleep
Obesity, snoring, hypertension	Polysomnogram or home sleep test	Obstructive sleep apnea (Chap. 297)	Continuous positive airway pressure; upper airway surgery (e.g., uvulopalatopharyngoplasty); dental appliance; weight loss
Cataplexy, hypnagogic hallucinations, sleep paralysis	Polysomnogram and multiple sleep latency test	Narcolepsy	Stimulants (e.g., modafinil, methylphenidate); REM sleep-suppressing antidepressants (e.g., venlafaxine); pitolisant; solriamfetol; sodium oxybate
Restless legs, kicking movements during sleep	Assessment for predisposing medical conditions (e.g., iron deficiency or renal failure)	Restless legs syndrome with or without periodic limb movements	Treatment of predisposing condition; dopamine agonists (e.g., pramipexole, ropinirole); gabapentin; pregabalin; opiates
Sedating medications, stimulant withdrawal, head trauma, systemic inflammation, Parkinson's disease and other neurodegenerative disorders, hypothyroidism, encephalopathy	Thorough medical history and examination including detailed neurologic examination	Sleepiness due to a drug or medical condition	Change medications, treat underlying condition, consider stimulants

Eliciting a history of daytime sleepiness is usually adequate, but objective quantification is sometimes necessary. The MSLT measures a patient's propensity to sleep under quiet conditions. An overnight polysomnogram should precede the MSLT to establish that the patient has had an adequate amount of good-quality nighttime sleep. The MSLT consists of five 20-min nap opportunities every 2 h across the day. The patient is instructed to try to fall asleep, and the major endpoints are the average latency to sleep and the occurrence of REM sleep during the naps. An average sleep latency across the naps of <8 min is considered objective evidence of excessive daytime sleepiness. REM sleep normally occurs only during nighttime sleep, and the occurrence of REM sleep in two or more of the MSLT daytime naps provides support for the diagnosis of narcolepsy.

For the safety of the individual and the general public, physicians have a responsibility to help manage issues around driving in patients with sleepiness. Legal reporting requirements vary between states and countries, but at a minimum, physicians should inform sleepy patients about their increased risk of having an accident and advise such patients not to drive a motor vehicle until the sleepiness has been treated effectively. This discussion is especially important for commercial drivers, and it should be documented in the patient's medical record.

■ INSUFFICIENT SLEEP

Insufficient sleep is probably the most common cause of excessive daytime sleepiness. The average adult needs 7.5–8 h of sleep, but on weeknights the average U.S. adult gets only 6.75 h of sleep. Only 30% of the U.S. adult population reports consistently obtaining sufficient sleep. Insufficient sleep is especially common among shift workers, individuals working multiple jobs, and people in lower socioeconomic groups. Most teenagers need ≥9 h of sleep, but many fail to get enough sleep because of circadian phase delay, plus social pressures to stay up late coupled with early school start times. Late evening light exposure, television viewing, video-gaming, social media, texting, and smartphone use often delay bedtimes, despite the fixed early wake times required for work or school. As is typical with any disorder that causes sleepiness, individuals with chronically insufficient sleep may feel inattentive, irritable, unmotivated, and depressed, and have difficulty with school, work, and driving. Individuals differ in their optimal amount of sleep, and it can be helpful to ask how much sleep the patient obtains on a quiet vacation when he or she can sleep without restrictions. Some patients may think that a short amount of sleep is normal or advantageous, and they may not appreciate their biological need for more sleep, especially if coffee and other stimulants mask the sleepiness. A 2-week sleep log documenting the timing of sleep and daily level of alertness is diagnostically useful and provides helpful feedback for the patient. Extending sleep to the optimal amount on a regular basis can resolve the sleepiness and other symptoms. As with any lifestyle change, extending sleep requires commitment and adjustments, but the improvements in daytime alertness make this change worthwhile.

■ SLEEP APNEA SYNDROMES

Respiratory dysfunction during sleep is a common, serious cause of excessive daytime sleepiness as well as of disturbed nocturnal sleep. At least 24% of middle-aged men and 9% of middle-aged women in the United States have a reduction or cessation of breathing dozens or more times each night during sleep, with 9% of men and 4% of women doing so more than a hundred times per night. These episodes may be due to an occlusion of the airway (*obstructive sleep apnea*), absence of respiratory effort (*central sleep apnea*), or a combination of these factors. Failure to recognize and treat these

conditions appropriately may reduce daytime alertness and increase the risk of sleep-related motor vehicle crashes, depression, hypertension, myocardial infarction, diabetes, stroke, and mortality. Sleep apnea is particularly prevalent in overweight men and in the elderly, yet it is estimated to go undiagnosed in most affected individuals. This is unfortunate because several effective treatments are available. Readers are referred to Chap. 297 for a comprehensive review of the diagnosis and treatment of patients with sleep apnea.

■ NARCOLEPSY

Narcolepsy is characterized by difficulty sustaining wakefulness, poor regulation of REM sleep, and disturbed nocturnal sleep. All patients with narcolepsy have excessive daytime sleepiness. This sleepiness is usually moderate to severe, and in contrast to patients with disrupted sleep (e.g., sleep apnea), people with narcolepsy usually feel well rested upon awakening and then feel tired throughout much of the day. They may fall asleep at inappropriate times, but then feel refreshed again after a nap. In addition, they often experience symptoms related to an intrusion of REM sleep characteristics into wakefulness. REM sleep is characterized by dreaming and muscle paralysis, and people with narcolepsy can have: (1) sudden muscle weakness without a loss of consciousness, which is usually triggered by strong emotions (cataplexy; [Video 31-1](#)); (2) dream-like hallucinations at sleep onset (hypnagogic hallucinations) or upon awakening (hypnopompic hallucinations); and (3) muscle paralysis upon awakening (sleep paralysis). With severe cataplexy, an individual may be laughing at a joke and then suddenly collapse to the ground, immobile but awake for 1–2 min. With milder episodes, patients may have partial weakness of the face or neck. Narcolepsy is one of the more common causes of chronic sleepiness and affects about 1 in 2000 people in the United States. Narcolepsy typically begins between age 10 and 20; once established, the disease persists for life.

Narcolepsy is caused by loss of the hypothalamic neurons that produce the orexin neuropeptides (also known as hypocretins). Research in mice and dogs first demonstrated that a loss of orexin signaling due to null mutations of either the orexin neuropeptides or one of the orexin receptors causes sleepiness and cataplexy nearly identical to that seen in people with narcolepsy. Although genetic mutations rarely cause human narcolepsy, researchers soon discovered that patients with narcolepsy with cataplexy (now called type 1 narcolepsy) have very low or undetectable levels of orexins in their CSF, and autopsy studies showed a nearly complete loss of the orexin-producing neurons in the hypothalamus. The orexins normally promote long episodes of wakefulness and suppress REM sleep, and thus loss of orexin signaling results in frequent intrusions of sleep during the usual waking episode, with REM sleep and fragments of REM sleep at any time of day ([Fig. 31-3](#)). Patients with narcolepsy but no cataplexy (type 2 narcolepsy) usually have normal orexin levels and may have other yet uncharacterized causes of their excessive daytime sleepiness.

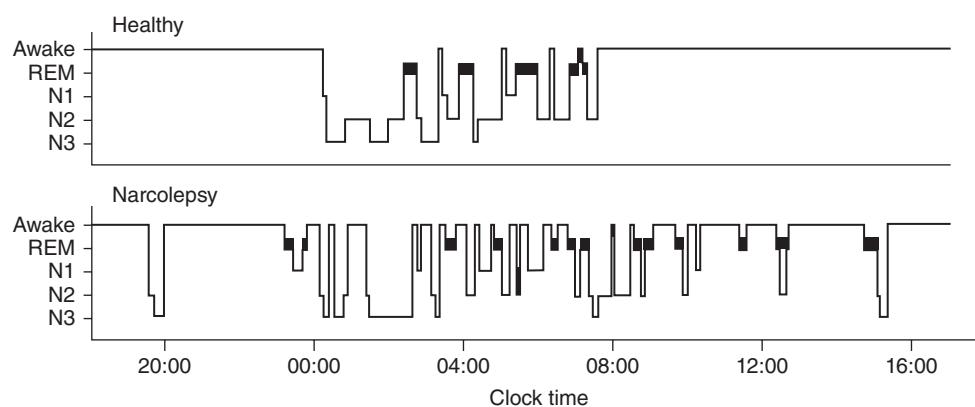


FIGURE 31-3 Polysomnographic recordings of a healthy individual and a patient with narcolepsy. The healthy individual has a long period of NREM sleep before entering REM sleep, but the individual with narcolepsy enters rapid eye movement (REM) sleep quickly at night and has moderately fragmented sleep. During the day, the healthy subject stays awake from 8:00 A.M. until midnight, but the patient with narcolepsy dozes off frequently, with many daytime naps that include REM sleep.

Extensive evidence suggests that an autoimmune process likely causes this selective loss of the orexin-producing neurons. Certain human leukocyte antigens (HLAs) can increase the risk of autoimmune disorders ([Chap. 350](#)), and narcolepsy has the strongest known HLA association. HLA DQB1*06:02 is found in >90% of people with type 1 narcolepsy, whereas it occurs in only 12–25% of the general population. Researchers now hypothesize that in people with DQB1*06:02, an immune response against influenza, *Streptococcus*, or other infections may also damage the orexin-producing neurons through a process of molecular mimicry. This mechanism may account for the eight- to twelvefold increase in new cases of narcolepsy among children in Europe who received a particular brand of H1N1 influenza A vaccine (Pandemrix). In support of this hypothesis, people with type 1 narcolepsy have heightened T cell responses against orexin peptides.

On rare occasions, narcolepsy can occur with neurologic disorders such as tumors or strokes that directly damage the orexin-producing neurons in the hypothalamus or their projections.

Diagnosis Narcolepsy is most commonly diagnosed by the history of chronic sleepiness plus cataplexy or other symptoms. Many disorders can cause feelings of weakness, but with true cataplexy patients will describe definite functional weakness (e.g., slurred speech, dropping a cup, slumping into a chair) that has consistent emotional triggers such as laughing at a joke, happy surprise at unexpectedly seeing a friend, or intense anger. Cataplexy occurs in about half of all narcolepsy patients and is diagnostically very helpful because it occurs in almost no other disorder. In contrast, occasional hypnagogic hallucinations and sleep paralysis occur in about 20% of the general population, and these symptoms are not as diagnostically specific.

When narcolepsy is suspected, the diagnosis should be firmly established with a polysomnogram followed the next day by an MSLT. The polysomnogram helps rule out other possible causes of sleepiness such as sleep apnea and establishes that the patient had adequate sleep the night before, and the MSLT provides essential, objective evidence of sleepiness plus REM sleep dysregulation. Across the five naps of the MSLT, most patients with narcolepsy will fall asleep in <8 min on average, and they will have episodes of REM sleep in at least two of the naps. Abnormal regulation of REM sleep is also manifested by the appearance of REM sleep within 15 min of sleep onset at night, which is rare in healthy individuals sleeping at their habitual bedtime. Stimulants should be stopped 1 week before the MSLT and antidepressants should be stopped 3 weeks prior, because these medications can affect the MSLT. In addition, patients should be encouraged to obtain a fully adequate amount of sleep each night for the week prior to the test to eliminate any effects of insufficient sleep.

TREATMENT

Narcolepsy

The treatment of narcolepsy is symptomatic. Most patients with narcolepsy feel more alert after sleep, and they should be encouraged to get adequate sleep each night and to take a 15- to 20-min nap in the afternoon. This nap may be sufficient for some patients with mild narcolepsy, but most also require treatment with wake-promoting medications. Modafinil is often used because it has fewer side effects than amphetamines and a relatively long half-life; for most patients, 200–400 mg each morning is very effective. Methylphenidate (10–20 mg bid) or dextroamphetamine (10 mg bid) are also effective, but sympathomimetic side effects, anxiety, and the potential for abuse can be concerns. These medications are available in slow-release formulations, extending their duration of action and allowing easier dosing. Solriamfetol, a norepinephrine-dopamine reuptake inhibitor (75–150 mg daily), and pitolisant, a selective histamine 3 (H₃) receptor antagonist (8.9–35.6 mg daily), also improve sleepiness and have relatively few side effects. Sodium oxybate (gamma hydroxybutyrate), given at bedtime and 3–4 h later, is often very valuable in improving alertness, but it can produce excessive sedation, nausea, and confusion.

Cataplexy is usually much improved with antidepressants that increase noradrenergic or serotonergic tone because these neurotransmitters strongly suppress REM sleep and cataplexy. Venlafaxine (37.5–150 mg each morning) and fluoxetine (10–40 mg each morning) are often quite effective. The tricyclic antidepressants, such as protriptyline (10–40 mg/d) or clomipramine (25–50 mg/d) are potent suppressors of cataplexy, but their anticholinergic effects, including sedation and dry mouth, make them less attractive.¹ Sodium oxybate, twice each night, is also very helpful in reducing cataplexy.

¹No antidepressant has been approved by the US Food and Drug Administration (FDA) for treating narcolepsy.

EVALUATION OF INSOMNIA

Insomnia is the complaint of poor sleep and usually presents as difficulty initiating or maintaining sleep. People with insomnia are dissatisfied with their sleep and feel that it impairs their ability to function well in work, school, and social situations. Affected individuals often experience fatigue, decreased mood, irritability, malaise, and cognitive impairment.

Chronic insomnia, lasting >3 months, occurs in about 10% of adults and is more common in women, older adults, people of lower socioeconomic status, and individuals with medical, psychiatric, and substance abuse disorders. Acute or short-term insomnia affects over 30% of adults and is often precipitated by stressful life events such as a major illness or loss, change of occupation, medications, and substance abuse. If the acute insomnia triggers maladaptive behaviors such as increased nocturnal light exposure, frequently checking the clock, or attempting to sleep more by napping, it can lead to chronic insomnia.

Most insomnia begins in adulthood, but many patients may be predisposed and report easily disturbed sleep predating the insomnia, suggesting that their sleep is lighter than usual. Clinical studies and animal models indicate that insomnia is associated with activation during sleep of brain areas normally active only during wakefulness. The polysomnogram is rarely used in the evaluation of insomnia, as it typically confirms the patient's subjective report of long latency to sleep and numerous awakenings but usually adds little new information. Many patients with insomnia have increased fast (beta) activity in the EEG during sleep; this fast activity is normally present only during wakefulness, which may explain why some patients report feeling awake for much of the night. The MSLT is rarely used in the evaluation of insomnia because, despite their feelings of low energy, most people with insomnia do not easily fall asleep during the day, and on the MSLT, their average sleep latencies are usually longer than normal.

Many factors can contribute to insomnia, and obtaining a careful history is essential so one can select therapies targeting the underlying factors. The assessment should focus on identifying predisposing, precipitating, and perpetuating factors.

Psychophysiological Factors Many patients with insomnia have negative expectations and conditioned arousal that interfere with sleep. These individuals may worry about their insomnia during the day and have increasing anxiety as bedtime approaches if they anticipate a poor night of sleep. While attempting to sleep, they may frequently check the clock, which only heightens anxiety and frustration. They may find it easier to sleep in a new environment rather than their bedroom, as it lacks the negative associations.

Inadequate Sleep Hygiene Patients with insomnia sometimes develop counterproductive behaviors that contribute to their insomnia. These can include daytime napping that reduces sleep drive at night; an irregular sleep-wake schedule that disrupts their circadian rhythms; use of wake-promoting substances (e.g., caffeine, tobacco) too close to bedtime; engaging in alerting or stressful activities close to bedtime (e.g., arguing with a partner, work-related emailing and texting while in bed, sleeping with a smartphone or tablet at the bedside); and routinely using the bedroom for activities other than sleep or sex (e.g., email,

television, work), so the bedroom becomes associated with arousing or stressful feelings.

Psychiatric Conditions About 80% of patients with psychiatric disorders have sleep complaints, and about half of all chronic insomnia occurs in association with a psychiatric disorder. Depression is classically associated with early morning awakening, but it can also interfere with the onset and maintenance of sleep. Mania and hypomania can disrupt sleep and often are associated with substantial reductions in the total amount of sleep. Anxiety disorders can lead to racing thoughts and rumination that interfere with sleep and can be very problematic if the patient's mind becomes active midway through the night. Panic attacks can arise from sleep and need to be distinguished from other parasomnias. Insomnia is common in schizophrenia and other psychoses, often resulting in fragmented sleep, less deep NREM sleep, and sometimes reversal of the day-night sleep pattern.

Medications and Drugs of Abuse A wide variety of psychoactive drugs can interfere with sleep. Caffeine, which has a half-life of 6–9 h, can disrupt sleep for up to 8–14 h, depending on the dose, variations in metabolism, and an individual's caffeine sensitivity. Insomnia can also result from use of prescription medications too close to bedtime (e.g., antidepressants, stimulants, glucocorticoids, theophylline). Conversely, withdrawal of sedating medications such as alcohol, narcotics, or benzodiazepines can cause insomnia. Alcohol taken just before bed can shorten sleep latency, but it often produces rebound insomnia 2–3 h later as it wears off. This same problem with sleep maintenance can occur with short-acting medications such as alprazolam or zolpidem.

Medical Conditions A large number of medical conditions disrupt sleep. Pain from rheumatologic disorders or a painful neuropathy commonly disrupts sleep. Some patients may sleep poorly because of respiratory conditions such as asthma, chronic obstructive pulmonary disease, cystic fibrosis, congestive heart failure, or restrictive lung disease, and some of these disorders are worse at night due to circadian variations in airway resistance and postural changes in bed that can result in nocturnal dyspnea. Many women experience poor sleep with the hormonal changes of menopause. Gastroesophageal reflux is also a common cause of difficulty sleeping.

Neurologic Disorders Dementia (Chap. 29) is often associated with poor sleep, probably due to a variety of factors, including napping during the day, altered circadian rhythms, and perhaps a weakened output of the brain's sleep-promoting mechanisms. In fact, insomnia and nighttime wandering are some of the most common causes for institutionalization of patients with dementia, because they place a larger burden on caregivers. Conversely, in cognitively intact elderly men, fragmented sleep and poor sleep quality are associated with subsequent cognitive decline. Patients with Parkinson's disease may sleep poorly due to rigidity, dementia, and other factors. Fatal familial insomnia is a very rare neurodegenerative condition caused by mutations in the prion protein gene (Chap. 438), and although insomnia is a common early symptom, most patients present with other obvious neurologic signs such as dementia, myoclonus, dysarthria, or autonomic dysfunction.

TREATMENT

Insomnia

Treatment of insomnia improves quality of life and can promote long-term health. With improved sleep, patients often report less daytime fatigue, improved cognition, and more energy. Treating the insomnia can also improve comorbid disease. For example, management of insomnia at the time of diagnosis of major depression often improves the response to antidepressants and reduces the risk of relapse. Sleep loss can heighten the perception of pain, so a similar approach is warranted in acute and chronic pain management.

The treatment plan should target all putative contributing factors: establish good sleep hygiene, treat medical disorders, use behavioral therapies for anxiety and negative conditioning, and use

pharmacotherapy and/or psychotherapy for psychiatric disorders. Behavioral therapies should be the first-line treatment, followed by judicious use of sleep-promoting medications if needed.

TREATMENT OF MEDICAL AND PSYCHIATRIC DISEASE

If the history suggests that a medical or psychiatric disease contributes to the insomnia, then it should be addressed by, for example, treating the pain or depression, improving breathing, and switching or adjusting the timing of medications.

IMPROVE SLEEP HYGIENE

Attention should be paid to improving sleep hygiene and avoiding counterproductive, arousing behaviors before bedtime. Patients should establish a regular bedtime and wake time, even on weekends, to help synchronize their circadian rhythms and sleep patterns. The amount of time allocated for sleep should not be more than their actual total amount of sleep. In the 30 min before bedtime, patients should establish a relaxing "wind-down" routine that can include a warm bath, listening to music, meditation, or other relaxation techniques. The bedroom should be off-limits to computers, televisions, radios, smartphones, videogames, and tablets. If an e-reader is used, the light should be adjusted for evening use (dimmer and reduced blue light) if possible, because light itself, especially in the blue spectrum, suppresses melatonin secretion and is arousing. Once in bed, patients should try to avoid thinking about anything stressful or arousing such as problems with relationships or work. If they cannot fall asleep within 20 min, it often helps to get out of bed and read or listen to relaxing music in dim light as a form of distraction from any anxiety, but artificial light, including light from a television, cell phone, or computer, should be avoided.

Table 31-2 outlines some of the key aspects of good sleep hygiene to improve insomnia.

COGNITIVE BEHAVIORAL THERAPY

Cognitive behavioral therapy (CBT) uses a combination of the techniques above plus additional methods to improve insomnia. A trained therapist may use cognitive psychology techniques to reduce excessive worrying about sleep and to reframe faulty beliefs about the insomnia and its daytime consequences. The therapist may also teach the patient relaxation techniques, such as progressive muscle relaxation or meditation, to reduce autonomic arousal, intrusive thoughts, and anxiety.

MEDICATIONS FOR INSOMNIA

If insomnia persists after treatment of these contributing factors, pharmacotherapy is often used on a nightly or intermittent basis. A variety of sedatives can improve sleep.

Antihistamines, such as diphenhydramine, are the primary active ingredient in most over-the-counter sleep aids. These may be of

TABLE 31-2 Methods to Improve Sleep Hygiene in Insomnia Patients

HELPFUL BEHAVIORS	BEHAVIORS TO AVOID
Use the bed only for sleep and sex • If you cannot sleep within 20 min, get out of bed and read or do other relaxing activities in dim light before returning to bed	Avoid behaviors that interfere with sleep physiology, including: • Napping, especially after 3:00 PM • Attempting to sleep too early • Caffeine after lunchtime
Make quality sleep a priority • Go to bed and get up at the same time each day • Ensure a restful environment (comfortable bed, bedroom quiet and dark)	In the 2–3 h before bedtime, avoid: • Heavy eating • Smoking or alcohol • Vigorous exercise
Develop a consistent bedtime routine. For example: • Prepare for sleep with 20–30 min of relaxation (e.g., soft music, meditation, yoga, pleasant reading) • Take a warm bath	When trying to fall asleep, avoid: • Solving problems • Thinking about life issues • Reviewing events of the day

benefit when used intermittently but can produce tolerance and anticholinergic side effects such as dry mouth and constipation, which limit their use, particularly in the elderly.

Benzodiazepine receptor agonists (BzRAs) are an effective and well-tolerated class of medications for insomnia. BzRAs bind to the GABA_A receptor and potentiate the postsynaptic response to GABA. GABA_A receptors are found throughout the brain, and BzRAs may globally reduce neural activity and enhance the activity of specific sleep-promoting GABAergic pathways. Classic BzRAs include lorazepam, triazolam, and clonazepam, whereas newer agents such as zolpidem and zaleplon have more selective affinity for the α_1 subunit of the GABA_A receptor.

Specific BzRAs are often chosen based on the desired duration of action. The most commonly prescribed agents in this family are zaleplon (5–20 mg), with a half-life of 1–2 h; zolpidem (5–10 mg) and triazolam (0.125–0.25 mg), with half-lives of 2–4 h; eszopiclone (1–3 mg), with a half-life of 5–8 h; and temazepam (15–30 mg), with a half-life of 8–20 h. Generally, side effects are minimal when the dose is kept low and the serum concentration is minimized during the waking hours (by using the shortest-acting effective agent). For chronic insomnia, intermittent use is recommended, unless the consequences of untreated insomnia outweigh concerns regarding chronic use.

The heterocyclic antidepressants (trazodone, amitriptyline,² and doxepin) are the most commonly prescribed alternatives to BzRAs due to their lack of abuse potential and low cost. Trazodone (25–100 mg) is used more commonly than the tricyclic antidepressants, because it has a much shorter half-life (5–9 h) and less anticholinergic activity.

The orexin receptor antagonists suvorexant (10–20 mg) and lemborexant (5–10 mg) can also improve insomnia by blocking the wake-promoting effects of the orexin neuropeptides. These have long half-lives and can produce morning sedation, and as they reduce orexin signaling, they can rarely produce hypnagogic hallucinations and sleep paralysis (see narcolepsy section above).

Medications for insomnia are now among the most commonly prescribed medications, but they should be used cautiously. All sedatives increase the risk of injurious falls and confusion in the elderly, and therefore if needed these medications should be used at the lowest effective dose. Morning sedation can interfere with driving and judgment, and when selecting a medication, one should consider the duration of action. Benzodiazepines carry a risk of addiction and abuse, especially in patients with a history of alcohol or sedative abuse. In patients with depression, all sedatives can worsen the depression. Like alcohol, some sleep-promoting medications can worsen sleep apnea. Sedatives can also produce complex behaviors during sleep, such as sleepwalking and sleep eating, especially at higher doses.

²Trazodone and amitriptyline have not been approved by the FDA for treating insomnia.

■ RESTLESS LEGS SYNDROME

Patients with restless legs syndrome (RLS) report an irresistible urge to move the legs. Many patients report a creepy-crawly or unpleasant deep ache within the thighs or calves, and those with more severe RLS may have discomfort in the arms as well. For most patients with RLS, these dysesthesias and restlessness are much worse in the evening and first half of the night. The symptoms appear with inactivity and can make sitting still in an airplane or when watching a movie a miserable experience. The sensations are temporarily relieved by movement, stretching, or massage. This nocturnal discomfort usually interferes with sleep, and patients may report daytime sleepiness as a consequence. RLS is very common, affecting 5–10% of adults, and is more common in women and older adults.

A variety of factors can cause RLS. Iron deficiency is the most common treatable cause, and iron replacement should be considered if the ferritin level is <75 ng/mL. RLS can also occur with peripheral

neuropathies and uremia and can be worsened by pregnancy, caffeine, alcohol, antidepressants, lithium, neuroleptics, and antihistamines. Genetic factors contribute to RLS, and polymorphisms in a variety of genes (*BTBD9*, *MEIS1*, *MAP2K5/LBXCOR*, and *PTPRD*) have been linked to RLS, although as yet, the mechanism through which they cause RLS remains unknown. Roughly one-third of patients (particularly those with an early age of onset) have multiple affected family members.

RLS is treated by addressing the underlying cause such as iron deficiency if present. Otherwise, treatment is symptomatic, and dopamine agonists or alpha-2-delta calcium channel ligands are used most frequently. Agonists of dopamine D_{2/3} receptors such as pramipexole (0.25–0.5 mg q7PM) or ropinirole (0.5–4 mg q7PM) are usually quite effective, but about 25% of patients taking dopamine agonists develop augmentation, a worsening of RLS such that symptoms begin earlier in the day and can spread to other body regions. Other possible side effects of dopamine agonists include nausea, morning sedation, and increases in rewarding behaviors such as sex and gambling. Alpha-2-delta calcium channel ligands such as gabapentin (300–600 mg q7PM) and pregabalin (150–450 mg q7PM) can also be quite effective; these are less likely to cause augmentation, and they can be especially helpful in patients with concomitant pain, neuropathy, or anxiety. Opioids and benzodiazepines may also be of therapeutic value. Most patients with restless legs also experience PLMD, although the reverse is not the case.

■ PERIODIC LIMB MOVEMENT DISORDER

PLMD involves rhythmic twitches of the legs that disrupt sleep. The movements resemble a triple flexion reflex with extensions of the great toe and dorsiflexion of the foot for 0.5–5.0 s, which recur every 20–40 s during NREM sleep, in episodes lasting from minutes to hours. PLMD is diagnosed by a polysomnogram that includes recordings of the anterior tibialis and sometimes other muscles. The EEG shows that the movements of PLMD frequently cause brief arousals that disrupt sleep and can cause insomnia and daytime sleepiness. PLMD can be caused by the same factors that cause RLS (see above), and the frequency of leg movements improves with the same medications used for RLS, including dopamine agonists. Genetic studies identified polymorphisms associated with both RLS and PLMD, suggesting that they may have a common pathophysiology.

■ PARASOMNIAS

Parasomnias are abnormal behaviors or experiences that arise from or occur during sleep. A variety of parasomnias can occur during NREM sleep, from brief confusional arousals to sleepwalking and night terrors. The presenting complaint is usually related to the behavior itself, but the parasomnias can disturb sleep continuity or lead to mild impairments in daytime alertness. Two main parasomnias occur in REM sleep: REM sleep behavior disorder (RBD) and nightmares.

Sleepwalking (Somnambulism) Patients affected by this disorder carry out automatic motor activities that range from simple to complex. Individuals may walk, urinate inappropriately, eat, exit the house, or drive a car with minimal awareness. It may be difficult to arouse the patient to wakefulness, and some individuals may respond to attempted awakening with agitation or violence. In general, it is safest to lead the patient back to bed, at which point he or she will often fall back asleep. Sleepwalking arises from NREM stage N3 sleep, usually in the first few hours of the night, and the EEG initially shows the slow cortical activity of deep NREM sleep even when the patient is moving about. Sleepwalking is most common in children and adolescents, when deep NREM sleep is most abundant. About 15% of children have occasional sleepwalking, and it persists in about 1% of adults. Episodes are usually isolated but may be recurrent in 1–6% of patients. The cause is unknown, although it has a familial basis in roughly one-third of cases. Sleepwalking can be worsened by stress, alcohol, and insufficient sleep, which subsequently causes an increase in deep NREM sleep. These should be addressed if present. Small studies have shown some efficacy of antidepressants and benzodiazepines;

relaxation techniques and hypnosis can also be helpful. Patients and their families should improve home safety (e.g., replace glass doors, remove low tables to avoid tripping) to minimize the chance of injury if sleepwalking occurs.

Sleep Terrors This disorder occurs primarily in young children during the first few hours of sleep during NREM stage N3 sleep. The child often sits up during sleep and screams, exhibiting autonomic arousal with sweating, tachycardia, large pupils, and hyperventilation. The individual may be difficult to arouse and rarely recalls the episode on awakening in the morning. Treatment usually consists of reassuring parents that the condition is self-limited and benign, and like sleepwalking, it may improve by avoiding insufficient sleep.

Sleep Enuresis Bedwetting, like sleepwalking and night terrors, is another parasomnia that occurs during sleep in the young. Before age 5 or 6 years, nocturnal enuresis should be considered a normal feature of development. The condition usually improves spontaneously by puberty, persists in 1–3% of adolescents, and is rare in adulthood. Treatment consists of bladder training exercises and behavioral therapy. Symptomatic pharmacotherapy is usually accomplished in adults with desmopressin (0.2 mg qhs), oxybutynin chloride (5 mg qhs), or imipramine (10–25 mg qhs). Important causes of nocturnal enuresis in patients who were previously continent for 6–12 months include urinary tract infections or malformations, cauda equina lesions, emotional disturbances, epilepsy, sleep apnea, and certain medications.

Sleep Bruxism Bruxism is an involuntary, forceful grinding of teeth during sleep that affects 10–20% of the population. The patient is usually unaware of the problem. The typical age of onset is 17–20 years, and spontaneous remission usually occurs by age 40. In many cases, the diagnosis is made during dental examination, damage is minor, and no treatment is indicated. In more severe cases, treatment with a mouth guard is necessary to prevent tooth injury. Stress management, benzodiazepines, and biofeedback can be useful when bruxism is a manifestation of psychological stress.

REM Sleep Behavior Disorder (RBD) RBD ([Video 31-2](#)) is distinct from other parasomnias in that it occurs during REM sleep. The patient or the bed partner usually reports agitated or violent behavior during sleep, and upon awakening, the patient can often report a dream that matches the accompanying movements. During normal REM sleep, nearly all nonrespiratory skeletal muscles are paralyzed, but in patients with RBD, dramatic limb movements such as punching or kicking lasting seconds to minutes occur during REM sleep, and it is not uncommon for the patient or the bed partner to be injured.

The prevalence of RBD increases with age, afflicting about 2% of adults aged >70, and is about twice as common in men. Within 12 years of disease onset, half of RBD patients develop a synucleinopathy such as Parkinson's disease ([Chap. 435](#)) or dementia with Lewy bodies ([Chap. 434](#)), or occasionally multiple system atrophy ([Chap. 440](#)), and over 90% develop a synucleinopathy by 25 years. RBD can occur in patients taking antidepressants, and in some, these medications may unmask this early indicator of neurodegeneration. Synucleinopathies probably cause neuronal loss in brainstem regions that regulate muscle paralysis during REM sleep, and loss of these neurons permits movements to break through during REM sleep. RBD also occurs in about 30% of patients with narcolepsy, but the underlying cause is probably different, as they seem to be at no increased risk of a neurodegenerative disorder.

Many patients with RBD have sustained improvement with clonazepam (0.5–2.0 mg qhs).³ Melatonin at doses up to 9 mg nightly may also prevent attacks.

CIRCADIAN RHYTHM SLEEP DISORDERS

A subset of patients presenting with either insomnia or hypersomnia may have a disorder of sleep *timing* rather than sleep *generation*.

Disorders of sleep timing can be either organic (i.e., due to an abnormality of circadian pacemaker[s]) or environmental/behavioral (i.e., due to a disruption of environmental synchronizers). Effective therapies aim to entrain the circadian rhythm of sleep propensity to the appropriate behavioral phase.

Delayed Sleep-Wake Phase Disorder DSWPD is characterized by: (1) sleep onset and wake times persistently later than desired; (2) actual sleep times at nearly the same clock hours daily; and (3) if conducted at the habitual delayed sleep time, essentially normal sleep on polysomnography (except for delayed sleep onset). About half of patients with DSWPD exhibit an abnormally delayed endogenous circadian phase, which can be assessed by measuring the onset of secretion of melatonin in either the blood or saliva; this is best done in a dimly lit environment as light suppresses melatonin secretion. Dim-light melatonin onset (DLMO) in DSWPD patients occurs later in the evening than normal, which is about 8:00–9:00 P.M. (i.e., about 1–2 h before habitual bedtime). Patients tend to be young adults. The delayed circadian phase could be due to: (1) an abnormally long, genetically determined intrinsic period of the endogenous circadian pacemaker; (2) reduced phase-advancing capacity of the pacemaker; (3) slower buildup of homeostatic sleep drive during wakefulness; or (4) an irregular prior sleep-wake schedule, characterized by frequent nights when the patient chooses to remain awake while exposed to artificial light well past midnight (for personal, social, school, or work reasons). In most cases, it is difficult to distinguish among these factors, as patients with either a behaviorally induced or biologically driven circadian phase delay may both exhibit a similar circadian phase delay in DLMO, and both factors make it difficult to fall asleep at the desired hour. Late onset of dim-light melatonin secretion can help distinguish DSWPD from other forms of sleep-onset insomnia. DSWPD is a chronic condition that can persist for years and may not respond to attempts to reestablish normal bedtime hours. Treatment methods involving phototherapy with blue-enriched light during the morning hours and/or melatonin administration in the evening hours show promise in these patients, although the relapse rate is high.

Advanced Sleep-Wake Phase Disorder Advanced sleep-wake phase disorder (ASWPD) is the converse of DSWPD. Most commonly, this syndrome occurs in older people, 15% of whom report that they cannot sleep past 5:00 A.M., with twice that number complaining that they wake up too early at least several times per week. Patients with ASWPD are sleepy during the evening hours, even in social settings. Sleep-wake timing in ASWPD patients can interfere with a normal social life. Patients with this circadian rhythm sleep disorder can be distinguished from those who have early wakening due to insomnia because ASWPD patients show early onset of dim-light melatonin secretion.

In addition to age-related ASWPD, an early-onset familial variant of this condition has also been reported. In two families in which ASWPD was inherited in an autosomal dominant pattern, the syndrome was due to missense mutations in a circadian clock component (in the casein kinase binding domain of PER2 in one family, and in casein kinase I delta in the other) that shortens the circadian period. Patients with ASWPD may benefit from bright light and/or blue enriched phototherapy during the evening hours to reset the circadian pacemaker to a later hour.

Non-24-h Sleep-Wake Rhythm Disorder Non-24-h sleep-wake rhythm disorder (N24SWD) most commonly occurs when the primary synchronizing input (i.e., the light-dark cycle) from the environment to the circadian pacemaker is lost (as occurs in many blind people with no light perception), and the maximal phase-advancing capacity of the circadian pacemaker in response to nonphotic cues cannot accommodate the difference between the 24-h geophysical day and the intrinsic period of the patient's circadian pacemaker, resulting in loss of entrainment to the 24-h day. The sleep of most blind patients with N24SWD is restricted to the nighttime hours due to social or occupational demands. Despite this regular sleep-wake schedule, affected patients with N24SWD are nonetheless unable to maintain

³No medications have been approved by the FDA for the treatment of RBD.

a stable phase relationship between the output of the non-entrained circadian pacemaker and the 24-h day. Therefore, most blind patients present with intermittent bouts of insomnia. When the blind patient's endogenous circadian rhythms are out of phase with the local environment, nighttime insomnia coexists with excessive daytime sleepiness. Conversely, when the endogenous circadian rhythms of those same patients are in phase with the local environment, symptoms remit. The interval between symptomatic phases may last several weeks to several months in blind patients with N24SWD, depending on the period of the underlying nonentrained rhythm and the 24-h day. Nightly low-dose (0.5 mg) melatonin administration may improve sleep and, in some cases, induce synchronization of the circadian pacemaker. In sighted patients, N24SWD can be caused by self-selected exposure to artificial light that inadvertently entrains the circadian pacemaker to a >24-h schedule, and these individuals present with an incremental pattern of successive delays in sleep timing, progressing in and out of phase with local time—a clinical presentation that is seldom seen in blind patients with N24SWD.

Shift-Work Disorder More than 7 million workers in the United States regularly work at night, either on a permanent or rotating schedule. Many more begin the commute to work or school between 4:00 A.M. and 7:00 A.M., requiring them to commute and then work during a time of day that they would otherwise be asleep. In addition, each week, millions of "day" workers and students elect to remain awake at night or awaken very early in the morning to work or study to meet work or school deadlines, drive long distances, compete in sporting events, or participate in recreational activities. Such schedules can result in both sleep loss and misalignment of circadian rhythms with respect to the sleep-wake cycle.

The circadian timing system usually fails to adapt successfully to the inverted schedules required by overnight work or the phase advance required by early morning (4:00 A.M. to 7:00 A.M.) start times. This leads to a misalignment between the desired work-rest schedule and the output of the pacemaker, resulting in disturbed daytime sleep in most such individuals. Excessive work hours (per day or per week), insufficient time off between consecutive days of work or school, and frequent travel across time zones may be contributing factors. Sleep deficiency, increased length of time awake prior to work, and misalignment of circadian phase impair alertness and performance, increase reaction time, and increase risk of performance lapses, thereby resulting in greater safety hazards among night workers and other sleep-deprived individuals. Sleep disturbance nearly doubles the risk of a fatal work accident. In addition, long-term night-shift workers have higher rates of breast, colorectal, and prostate cancer and of cardiac, gastrointestinal, metabolic, and reproductive disorders. The World Health Organization has added night-shift work to its list of probable carcinogens.

Sleep onset begins in local brain regions before gradually sweeping over the entire brain as sensory thresholds rise and consciousness is lost. A sleepy individual struggling to remain awake may attempt to continue performing routine and familiar motor tasks during the transition state between wakefulness and stage N1 sleep, while unable to adequately process sensory input from the environment. Such sleep-related attentional failures typically last only seconds but are known on occasion to persist for longer durations. Motor vehicle operators who fail to heed the warning signs of sleepiness are especially vulnerable to sleep-related accidents, as sleep processes can slow reaction times, induce automatic behavior, and intrude involuntarily upon the waking brain, causing catastrophic consequences—including 6400 fatalities and 50,000 debilitating injuries in the United States annually. For this reason, an expert consensus panel has concluded that individuals who have slept <2 h in the prior 24 h are unfit to drive a motor vehicle. There is a significant increase in the risk of sleep-related, fatal-to-the-driver highway crashes in the early morning and late afternoon hours, coincident with bimodal peaks in the daily rhythm of sleep tendency.

Physicians who work prolonged shifts, especially intermittent overnight shifts, constitute another group of workers at greater risk for accidents and other adverse consequences of lack of sleep and

misalignment of the circadian rhythm. Recurrent scheduling of resident physicians to work shifts of ≥24 consecutive hours impairs psychomotor performance to a degree that is comparable to alcohol intoxication, doubles the risk of attentional failures among intensive care unit resident physicians working at night, and significantly increases the risk of serious medical errors in intensive care units, including a fivefold increase in the risk of serious diagnostic mistakes. Some 20% of hospital resident physicians report making a fatigue-related mistake that injured a patient, and 5% admit making a fatigue-related mistake that resulted in the death of a patient. Moreover, working for >24 consecutive hours increases the risk of percutaneous injuries and more than doubles the risk of motor vehicle crashes during the commute home. For these reasons, in 2008, the National Academy of Medicine concluded that the practice of scheduling resident physicians to work for >16 consecutive hours without sleep is hazardous for both resident physicians and their patients.

Of individuals scheduled to work at night or in the early morning hours, 5–15% have much greater-than-average difficulties remaining awake during night work and sleeping during the day; these individuals are diagnosed with chronic and severe shift-work disorder (SWD). Patients with this disorder have a level of excessive sleepiness during work at night or in the early morning and insomnia during day sleep that the physician judges to be clinically significant; the condition is associated with an increased risk of sleep-related accidents and with some of the illnesses associated with night-shift work. Patients with chronic and severe SWD are profoundly sleepy at work. In fact, their sleep latencies during night work average just 2 min, comparable to mean daytime sleep latency durations of patients with narcolepsy or severe sleep apnea.

TREATMENT

Shift-Work Disorder

Caffeine is frequently used by night workers to promote wakefulness. However, it cannot forestall sleep indefinitely, and it does not shield users from sleep-related performance lapses. Postural changes, exercise, and strategic placement of nap opportunities can sometimes temporarily reduce the risk of fatigue-related performance lapses. Properly timed exposure to blue-enriched light or bright white light can directly enhance alertness and facilitate more rapid adaptation to night-shift work.

Modafinil (200 mg) or armodafinil (150 mg) 30–60 min before the start of an 8-h overnight shift is an effective treatment for the excessive sleepiness during night work in patients with SWD. Although treatment with modafinil or armodafinil significantly improves performance and reduces sleep propensity and the risk of lapses of attention during night work, affected patients remain excessively sleepy.

Fatigue risk management programs for night-shift workers should promote education about sleep, increase awareness of the hazards associated with sleep deficiency and night work, and screen for common sleep disorders. Work schedules should be designed to minimize: (1) exposure to night work; (2) the frequency of shift rotations; (3) the number of consecutive night shifts; and (4) the duration of night shifts.

Jet Lag Disorder Each year, >60 million people fly from one time zone to another, often resulting in excessive daytime sleepiness, sleep-onset insomnia, and frequent arousals from sleep, particularly in the latter half of the night. The syndrome is transient, typically lasting 2–14 d depending on the number of time zones crossed, the direction of travel, and the traveler's age and phase-shifting capacity. Travelers who spend more time outdoors at their destination reportedly adapt more quickly than those who remain in hotel or seminar rooms, presumably due to brighter (outdoor) light exposure. Avoidance of antecedent sleep loss or napping on the afternoon prior to overnight travel can reduce the difficulties associated with extended wakefulness. Laboratory studies suggest that low doses of melatonin can enhance

sleep efficiency, but only if taken when endogenous melatonin concentrations are low (i.e., during the biologic daytime).

In addition to jet lag associated with travel across time zones, many patients report a behavioral pattern that has been termed *social jet lag*, in which bedtimes and wake times on weekends or days off occur 4–8 h later than during the week. Such recurrent displacement of the timing of the sleep-wake cycle is common in adolescents and young adults and is associated with delayed circadian phase, sleep-onset insomnia, excessive daytime sleepiness, poorer academic performance, and increased risk of both obesity and depressive symptoms.

MEDICAL IMPLICATIONS OF CIRCADIAN RHYTHMICITY

Prominent circadian variations have been reported in the incidence of acute myocardial infarction, sudden cardiac death, and stroke, the leading causes of death in the United States. Platelet aggregability is increased in the early morning hours, coincident with the peak incidence of these cardiovascular events. Recurrent circadian disruption combined with chronic sleep deficiency, such as occurs during night-shift work, is associated with increased plasma glucose concentrations after a meal due to inadequate pancreatic insulin secretion. Night-shift workers with elevated fasting glucose have an increased risk of progressing to diabetes. Blood pressure of night workers with sleep apnea is higher than that of day workers. A better understanding of the possible role of circadian rhythmicity in the acute destabilization of a chronic condition such as atherosclerotic disease could improve the understanding of its pathophysiology.

Diagnostic and therapeutic procedures may also be affected by the time of day at which data are collected. Examples include blood pressure, body temperature, the dexamethasone suppression test, and plasma cortisol levels. The timing of chemotherapy administration has been reported to have an effect on the outcome of treatment. In addition, both the toxicity and effectiveness of drugs can vary with time of day. For example, more than a fivefold difference has been observed in mortality rates after administration of toxic agents to experimental animals at different times of day. Anesthetic agents are particularly sensitive to time-of-day effects. Finally, the physician must be aware of the public health risks associated with the ever-increasing demands made by the 24/7 schedules in our round-the-clock society.

ACKNOWLEDGMENT

John W. Winkelmann, MD, PhD, and Gary S. Richardson, MD, contributed to this chapter in prior editions, and some material from their work has been retained here.

FURTHER READING

- CASH RE et al: Association between sleep duration and ideal cardiovascular health among US adults, National Health and Nutrition Examination Survey. *Prev Chronic Dis* 17:E43, 2020.
- CHINOY ED et al: Unrestricted evening use of light-emitting tablet computers delays self-selected bedtime and disrupts circadian timing and alertness. *Physiol Rep* 6:e13692, 2018.
- FULTZ NE et al: Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep. *Science* 366:628, 2019.
- HOLTH JK et al: The sleep-wake cycle regulates brain interstitial fluid tau in mice and CSF tau in humans. *Science* 363:880, 2019.
- LANDRIGAN CP et al: Effect on patient safety of a resident physician schedule without 24-hour shifts. *N Engl J Med* 382:2514, 2020.
- LEE ML et al: High risk of near-crash driving events following night-shift work. *Proc Natl Acad Sci USA* 113:176, 2016.
- LIM AS et al: Sleep is related to neuron numbers in the ventrolateral preoptic/intermediate nucleus in older adults with and without Alzheimer's disease. *Brain* 137:2847, 2014.
- MCALPINE CS et al: Sleep modulates haematopoiesis and protects against atherosclerosis. *Nature* 566:383, 2019.
- RIEMANN D et al: The neurobiology, investigation, and treatment of chronic insomnia. *Lancet Neurol* 14:547, 2015.
- SCAMMELL TE: Narcolepsy. *N Engl J Med* 373:2654, 2015.
- SCAMMELL TE et al: Neural circuitry of wakefulness and sleep. *Neuron* 93:747, 2017.

SLETTEN TL et al: Efficacy of melatonin with behavioural sleep-wake scheduling for delayed sleep-wake phase disorder: a double-blind, randomised clinical trial. *PLoS Med* 15:e1002587, 2018.

VIDEO 31-1 A typical episode of severe cataplexy. The patient is joking and then falls to the ground with an abrupt loss of muscle tone. The electromyogram recordings (*four lower traces on the right*) show reductions in muscle activity during the period of paralysis. The electroencephalogram (*top two traces*) shows wakefulness throughout the episode. (*Video courtesy of Giuseppe Plazzi, University of Bologna.*)

VIDEO 31-2 Typical aggressive movements in rapid eye movement (REM) sleep behavior disorder. (*Video courtesy of Dr. Carlos Schenck, University of Minnesota Medical School.*)

Section 4 Disorders of Eyes, Ears, Nose, and Throat

32

Disorders of the Eye

Jonathan C. Horton



THE HUMAN VISUAL SYSTEM

The visual system provides a supremely efficient means for the rapid assimilation of information from the environment to aid in the guidance of behavior. The act of seeing begins with the capture of images focused by the cornea and lens on a light-sensitive membrane in the back of the eye called the *retina*. The retina is actually part of the brain, banished to the periphery to serve as a transducer for the conversion of patterns of light energy into neuronal signals. Light is absorbed by pigment in two types of photoreceptors: rods and cones. In the human retina, there are 100 million rods and 5 million cones. The rods operate in dim (scotopic) illumination. The cones function under daylight (photopic) conditions. The cone system is specialized for color perception and high spatial resolution. The majority of cones are within the macula, the portion of the retina that serves the central 10° of vision. In the middle of the macula, a small pit termed the *fovea*, packed exclusively with cones, provides the best visual acuity.

Photoreceptors hyperpolarize in response to light, activating bipolar, amacrine, and horizontal cells in the inner nuclear layer. After processing of photoreceptor responses by this complex retinal circuit, the flow of sensory information ultimately converges on a final common pathway: the ganglion cells. These cells translate the visual image impinging on the retina into a continuously varying barrage of action potentials that propagates along the primary optic pathway to visual centers within the brain. There are a million ganglion cells in each retina and hence a million fibers in each optic nerve.

Ganglion cell axons sweep along the inner surface of the retina in the nerve fiber layer, exit the eye at the optic disc, and travel through the optic nerve, optic chiasm, and optic tract to reach targets in the brain. The majority of fibers synapse on cells in the lateral geniculate body, a thalamic relay station. Cells in the lateral geniculate body project in turn to the primary visual cortex. This afferent retinogeniculocortical sensory pathway provides the neural substrate for visual perception. Although the lateral geniculate body is the main target of the retina, separate classes of ganglion cells project to other subcortical visual nuclei involved in different functions. Ganglion cells that mediate pupillary constriction and circadian rhythms are light sensitive owing to a novel visual pigment, melanopsin. Pupil responses are mediated by input to the pretectal olfactory nuclei in the midbrain. The pretectal nuclei send their output to the Edinger-Westphal nuclei, which in turn provide parasympathetic innervation to the iris sphincter via an interneuron in the ciliary ganglion. Circadian rhythms are

timed by a retinal projection to the suprachiasmatic nucleus. Visual orientation and eye movements are served by retinal input to the superior colliculus. Gaze stabilization and optokinetic reflexes are governed by a group of small retinal targets known collectively as the *brainstem accessory optic system*.

The eyes must be rotated constantly within their orbits to place and maintain targets of visual interest on the fovea. This activity, called *foveation*, or looking, is governed by an elaborate efferent motor system. Each eye is moved by six extraocular muscles that are supplied by cranial nerves from the oculomotor (III), trochlear (IV), and abducens (VI) nuclei. Activity in these ocular motor nuclei is coordinated by pontine and midbrain mechanisms for smooth pursuit, saccades, and gaze stabilization during head and body movements. Large regions of the frontal and parietooccipital cortex control these brainstem eye movement centers by providing descending supranuclear input.

CLINICAL ASSESSMENT OF VISUAL FUNCTION

REFRACTIVE STATE

In approaching a patient with reduced vision, the first step is to decide whether refractive error is responsible. In *emmetropia*, parallel rays from infinity are focused perfectly on the retina. Sadly, this condition is enjoyed by only a minority of the population. In *myopia*, the globe is too long, and light rays come to a focal point in front of the retina. Near objects can be seen clearly, but distant objects require a diverging lens in front of the eye. In *hyperopia*, the globe is too short, and hence, a converging lens is used to supplement the refractive power of the eye. In *astigmatism*, the corneal surface is not perfectly spherical, necessitating a cylindrical corrective lens. Most patients elect to wear eyeglasses or contact lenses to neutralize refractive error. An alternative is to permanently alter the refractive properties of the cornea by performing laser *in situ* keratomileusis (LASIK) or photorefractive keratectomy (PRK).

With the onset of middle age, *presbyopia* develops as the lens within the eye becomes unable to increase its refractive power to accommodate on near objects. To compensate for presbyopia, an emmetropic patient must use reading glasses. A patient already wearing glasses for distance correction usually switches to bifocals. The only exception is a myopic patient, who may achieve clear vision at near simply by removing glasses containing the distance prescription.

Refractive errors usually develop slowly and remain stable after adolescence, except in unusual circumstances. For example, the acute onset of diabetes mellitus can produce sudden myopia because of lens edema induced by hyperglycemia. Testing vision through a pinhole aperture is a useful way to screen quickly for refractive error. If visual acuity is better through a pinhole than it is with the unaided eye, the patient needs refraction to obtain best corrected visual acuity.

VISUAL ACUITY

The Snellen chart is used to test acuity at a distance of 6 m (20 ft). For convenience, a scale version of the Snellen chart called the Rosenbaum card is held at 36 cm (14 in.) from the patient (Fig. 32-1). All subjects should be able to read the 6/6 m (20/20 ft) line with each eye using their refractive correction, if any. Patients who need reading glasses because of presbyopia must wear them for accurate testing with the Rosenbaum card. If 6/6 (20/20) acuity is not present in each eye, the deficiency in vision must be explained. If it is worse than 6/240 (20/800), acuity should be recorded in terms of counting fingers, hand motions, light perception, or no light perception. Legal blindness is defined by the Internal Revenue Service as a best corrected acuity of 6/60 (20/200) or less in the better eye or a binocular visual field subtending 20° or less. Loss of vision in one eye only does not constitute legal blindness. For driving, the laws vary by state, but most require a corrected acuity of 6/12 (20/40) in at least one eye for unrestricted privileges. Patients who develop a homonymous hemianopia should not drive.

PUPILS

The pupils should be tested individually in dim light with the patient fixating on a distant target. There is no need to check the near response

ROSENBAUM POCKET VISION SCREENER



PUPIL GAUGE (mm.)

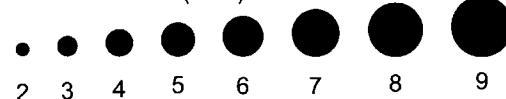


FIGURE 32-1 The Rosenbaum card is a miniature, scale version of the Snellen chart for testing visual acuity at near. When the visual acuity is recorded, the Snellen distance equivalent should bear a notation indicating that vision was tested at near, not at 6 m (20 ft), or else the Jaeger number system should be used to report the acuity. (Design Courtesy J.G. Rosenbaum MD.)

if the pupils respond briskly to light, because isolated loss of constriction (miosis) to accommodation does not occur. For this reason, the ubiquitous abbreviation PERRLA (pupils equal, round, and reactive to light and accommodation) implies a wasted effort with the last step. However, it is important to test the near response if the light response is poor or absent. Light-near dissociation occurs with neurosyphilis (Argyll Robertson pupil), with lesions of the dorsal midbrain (*Parinaud's syndrome*), and after aberrant regeneration (oculomotor nerve palsy, Adie's tonic pupil).

An eye with no light perception has no pupillary response to direct light stimulation. If the retina or optic nerve is only partially injured, the direct pupillary response will be weaker than the consensual pupillary response evoked by shining a light into the healthy fellow eye. A *relative afferent pupillary defect* (Marcus Gunn pupil) is elicited with the swinging flashlight test (Fig. 32-2). It is an extremely useful sign in retrobulbar optic neuritis and other optic nerve diseases, in which it may be the sole objective evidence for disease. In bilateral optic neuropathy, no afferent pupil defect is present if the optic nerves are affected equally.

Subtle inequality in pupil size, up to 0.5 mm, is a fairly common finding in normal persons. The diagnosis of essential or physiologic anisocoria is secure as long as the relative pupil asymmetry remains constant as ambient lighting varies. Anisocoria that increases in dim light indicates a sympathetic paresis of the iris dilator muscle. The triad of miosis with ipsilateral ptosis and anhidrosis constitutes *Horner's*

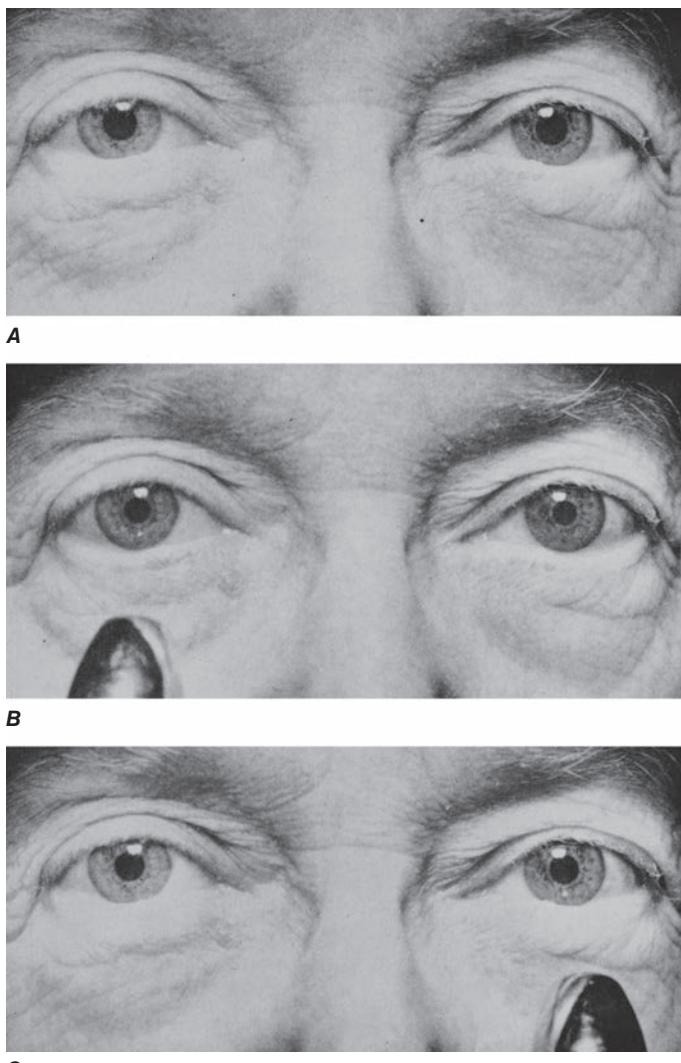


FIGURE 32-2 Demonstration of a relative afferent pupillary defect (Marcus Gunn pupil) in the left eye, done with the patient fixating on a distant target. *A*, With dim background lighting, the pupils are equal and relatively large. *B*, Shining a flashlight into the right eye evokes equal, strong constriction of both pupils. *C*, Swinging the flashlight over to the damaged left eye causes dilation of both pupils, although they remain smaller than in *A*. Swinging the flashlight back over to the healthy right eye would result in symmetric constriction back to the appearance shown in *B*. Note that the pupils always remain equal; the damage to the left retina/optic nerve is revealed by weaker bilateral pupil constriction to a flashlight in the left eye compared with the right eye. (From P Levatin: Arch Ophthalmol 62:768, 1959. Copyright © 1959 American Medical Association. All rights reserved.)

syndrome, although anhidrosis is an inconstant feature. A drop of 1% apraclonidine produces no effect on the normal pupil, but the miotic pupil dilates because of denervation hypersensitivity. Brainstem stroke, carotid dissection, and neoplasm impinging on the sympathetic chain occasionally are identified as the cause of Horner's syndrome, but most cases are idiopathic.

Anisocoria that increases in bright light suggests a parasympathetic palsy. The first concern is an oculomotor nerve paresis. This possibility is excluded if the eye movements are full and the patient has no ptosis or diplopia. Acute pupillary dilation (mydriasis) can result from damage to the ciliary ganglion in the orbit. Common mechanisms are infection (herpes zoster, influenza), trauma (blunt, penetrating, surgical), and ischemia (diabetes, temporal arteritis). After denervation of the iris sphincter, the pupil does not respond well to light, but the response to near is often relatively intact. When the near stimulus is removed, the pupil redilates very slowly compared with the normal pupil, hence the term *tonic pupil*. In *Adie's syndrome*, a tonic pupil is present, sometimes in conjunction with weak or absent tendon reflexes in the lower

extremities. This benign disorder, which occurs predominantly in healthy young women, is assumed to represent a mild dysautonomia. Tonic pupils are also associated with multiple system atrophy, segmental hypohidrosis, diabetes, and amyloidosis. Occasionally, a tonic pupil is discovered incidentally in an otherwise completely normal, asymptomatic individual. The diagnosis is confirmed by placing a drop of dilute (0.125%) pilocarpine into each eye. Denervation hypersensitivity produces pupillary constriction in a tonic pupil, whereas the normal pupil shows no response. Pharmacologic dilatation from accidental or deliberate instillation of anticholinergic (atropine, scopolamine) drops can produce pupillary mydriasis. Gardner's pupil refers to mydriasis induced by exposure to tropane alkaloids, contained in plants such as deadly nightshade, jimsonweed, or angel's trumpet. When an anticholinergic agent is responsible for pupil dilation, 1% pilocarpine causes no constriction.

Both pupils are affected equally by systemic medications. They are small with narcotic use (morphine, oxycodone) and large with anticholinergics (scopolamine). Parasympathetic agents (pilocarpine) used to treat glaucoma produce miosis. In any patient with an unexplained pupillary abnormality, a slit-lamp examination is helpful to exclude surgical trauma to the iris, an occult foreign body, perforating injury, intraocular inflammation, adhesions (synechia), angle-closure glaucoma, and iris sphincter rupture from blunt trauma.

EYE MOVEMENTS AND ALIGNMENT

Eye movements are tested by asking the patient, with both eyes open, to pursue a small target such as a pen tip into the cardinal fields of gaze. Normal ocular versions are smooth, symmetric, full, and maintained in all directions without nystagmus. Saccades, or quick refixation eye movements, are assessed by having the patient look back and forth between two stationary targets. The eyes should move rapidly and accurately in a single jump to their target. Ocular alignment can be judged by holding a penlight directly in front of the patient at about 1 m. If the eyes are straight, the corneal light reflex will be centered in the middle of each pupil. To test eye alignment more precisely, the cover test is useful. The patient is instructed to look at a small fixation target in the distance. One eye is occluded with a paddle or hand, while the other eye is observed. If the viewing eye shifts position to take up fixation on the target, it was misaligned. If it remains motionless, the first eye is uncovered and the test is repeated on the second eye. If neither eye moves, the eyes are aligned orthotropically. If the eyes are orthotropic in primary gaze but the patient complains of diplopia, the cover test should be performed with the head tilted or turned in whatever direction elicits diplopia. With practice, the examiner can detect an ocular deviation (heterotropia) as small as 1–2° with the cover test. In a patient with vertical diplopia, a small deviation can be difficult to detect and easy to dismiss. The magnitude of the deviation can be measured by placing a prism in front of the misaligned eye to determine the power required to neutralize the fixation shift evoked by covering the other eye. Temporary press-on plastic Fresnel prisms, prism eyeglasses, or eye muscle surgery can be used to restore binocular alignment.

STEREOPSIS

Stereoacluity is determined by presenting targets with retinal disparity separately to each eye by using polarized images. The most popular office tests measure a range of thresholds from 800 to 40 s of arc. Normal stereoacluity is 40 s of arc. If a patient achieves this level of stereoacluity, one is assured that the eyes are aligned orthotropically and that vision is intact in each eye. Random dot stereograms have no monocular depth cues and provide an excellent screening test for strabismus.

COLORVISION

The retina contains three classes of cones, with visual pigments of differing peak spectral sensitivity: red (560 nm), green (530 nm), and blue (430 nm). The red and green cone pigments are encoded on the X chromosome, and the blue cone pigment on chromosome 7. Mutations of the blue cone pigment are exceedingly rare. Mutations of the red and green pigments cause congenital X-linked color blindness in 8% of males. Affected individuals are not truly color blind; rather, they differ

from normal subjects in the way they perceive color and how they combine primary monochromatic lights to match a particular color. Anomalous trichromats have three cone types, but a mutation in one cone pigment (usually red or green) causes a shift in peak spectral sensitivity, altering the proportion of primary colors required to achieve a color match. Dichromats have only two cone types and therefore will accept a color match based on only two primary colors. Anomalous trichromats and dichromats have 6/6 (20/20) visual acuity, but their hue discrimination is impaired. Ishihara color plates can be used to detect red-green color blindness. The test plates contain a hidden number that is visible only to subjects with color confusion from red-green color blindness. Because color blindness is almost exclusively X-linked, it is worthwhile screening only male children.

The Ishihara plates often are used to detect acquired defects in color vision, although they are intended as a screening test for congenital color blindness. Acquired defects in color vision frequently result from disease of the macula or optic nerve. For example, patients with a history of optic neuritis often complain of color desaturation long after their visual acuity has returned to normal. Color blindness also can result from bilateral strokes involving the ventral portion of the occipital lobe (cerebral achromatopsia). Such patients can perceive only shades of gray and also may have difficulty recognizing faces (prosopagnosia) (Chap. 30). Infarcts of the dominant occipital lobe sometimes give rise to color anomia. Affected patients can discriminate colors but cannot name them.

VISUAL FIELDS

Vision can be impaired by damage to the visual system anywhere from the eyes to the occipital lobes. One can localize the site of the lesion with considerable accuracy by mapping the visual field deficit by finger confrontation and then correlating it with the topographic anatomy of the visual pathway (Fig. 32-3). Quantitative visual field mapping is performed by computer-driven perimeters that present a target of variable intensity at fixed positions in the visual field (Fig. 32-3A). By generating an automated printout of light thresholds, these static perimeters provide a sensitive means of detecting scotomas in the visual field. They are exceedingly useful for serial assessment of visual function in chronic diseases such as glaucoma and pseudotumor cerebri.

The crux of visual field analysis is to decide whether a lesion is before, at, or behind the optic chiasm. If a scotoma is confined to one eye, it must be due to a lesion anterior to the chiasm, involving either the optic nerve or the retina. Retinal lesions produce scotomas that correspond optically to their location in the fundus. For example, a superior-nasal retinal detachment results in an inferior-temporal field cut. Damage to the macula causes a central scotoma (Fig. 32-3B).

Optic nerve disease produces characteristic patterns of visual field loss. Glaucoma selectively destroys axons that enter the superotemporal or inferotemporal poles of the optic disc, resulting in arcuate scotomas shaped like a Turkish scimitar, which emanate from the blind spot and curve around fixation to end flat against the horizontal meridian (Fig. 32-3C). This type of field defect mirrors the arrangement of the nerve fiber layer in the temporal retina. Arcuate or nerve fiber layer scotomas also result from optic neuritis, ischemic optic neuropathy, optic disc drusen, and branch retinal artery or vein occlusion.

Damage to the entire upper or lower pole of the optic disc causes an altitudinal field cut that follows the horizontal meridian (Fig. 32-3D). This pattern of visual field loss is typical of ischemic optic neuropathy but also results from retinal vascular occlusion, advanced glaucoma, and optic neuritis.

About half the fibers in the optic nerve originate from ganglion cells serving the macula. Damage to papillomacular fibers causes a cecocentral scotoma that encompasses the blind spot and macula (Fig. 32-3E). If the damage is irreversible, pallor eventually appears in the temporal portion of the optic disc. Temporal pallor from a cecocentral scotoma may develop in optic neuritis, nutritional optic neuropathy, toxic optic neuropathy, Leber's hereditary optic neuropathy, Kjer's dominant optic atrophy, and compressive optic neuropathy. It is worth mentioning that the temporal side of the optic disc is slightly paler than the nasal side in most normal individuals. Therefore, it sometimes can be difficult

to decide whether the temporal pallor visible on fundus examination represents a pathologic change. Pallor of the nasal rim of the optic disc is a less equivocal sign of optic atrophy.

At the optic chiasm, fibers from nasal ganglion cells decussate into the contralateral optic tract. Crossed fibers are damaged more by compression than are uncrossed fibers. As a result, mass lesions of the sellar region cause a temporal hemianopia in each eye. Tumors anterior to the optic chiasm, such as meningiomas of the tuberculum sella, produce a junctional scotoma characterized by an optic neuropathy in one eye and a superior-temporal field cut in the other eye (Fig. 32-3G). More symmetric compression of the optic chiasm by a pituitary adenoma (see Fig. 380-1), meningioma, craniopharyngioma, glioma, or aneurysm results in a bitemporal hemianopia (Fig. 32-3H). The insidious development of a bitemporal hemianopia often goes unnoticed by the patient and will escape detection by the physician unless each eye is tested separately.

It is difficult to localize a postchiasmal lesion accurately, because injury anywhere in the optic tract, lateral geniculate body, optic radiations, or visual cortex can produce a homonymous hemianopia (i.e., a temporal hemifield defect in the contralateral eye and a matching nasal hemifield defect in the ipsilateral eye) (Fig. 32-3I). A unilateral postchiasmal lesion leaves the visual acuity in each eye unaffected, although the patient may read the letters on only the left or right half of the eye chart. Lesions of the optic radiations tend to cause poorly matched or incongruous field defects in each eye. Damage to the optic radiations in the temporal lobe (Meyer's loop) produces a superior quadrantic homonymous hemianopia (Fig. 32-3J), whereas injury to the optic radiations in the parietal lobe results in an inferior quadrantic homonymous hemianopia (Fig. 32-3K). Lesions of the primary visual cortex give rise to dense, congruous hemianopic field defects. Occlusion of the posterior cerebral artery supplying the occipital lobe is a common cause of total homonymous hemianopia. Some patients have macular sparing, because the central field representation at the tip of the occipital lobe is supplied by collaterals from the middle cerebral artery (Fig. 32-3L). Destruction of both occipital lobes produces cortical blindness. This condition can be distinguished from bilateral prechiasmal visual loss by noting that the pupil responses and optic fundi remain normal.

Partial recovery of homonymous hemianopia has been reported through computer-based rehabilitation therapy. During daily training sessions, patients fixate a central target while visual stimuli are presented within the blind region. The premise of vision restoration programs is that extra stimulation can promote recovery of partially damaged tissue located at the fringe of a cortical lesion. When fixation is controlled rigorously, however, no improvement of the visual fields can be demonstrated. No effective treatment exists for homonymous hemianopia caused by permanent brain damage.

DISORDERS

RED OR PAINFUL EYE

Corneal Abrasions Corneal abrasions are seen best by placing a drop of fluorescein in the eye and looking with the slit lamp, using a cobalt-blue light. A penlight with a blue filter will suffice if a slit lamp is not available. Damage to the corneal epithelium is revealed by yellow fluorescence of the basement membrane exposed by loss of the overlying epithelium. It is important to check for foreign bodies. To search the conjunctival fornices, the lower lid should be pulled down and the upper lid everted. A foreign body can be removed with a moistened cotton-tipped applicator after a drop of a topical anesthetic such as proparacaine has been placed in the eye. Alternatively, it may be possible to flush the foreign body from the eye by irrigating copiously with saline or artificial tears. If the corneal epithelium has been abraded, antibiotic ointment and a patch may be applied to the eye. A drop of an intermediate-acting cycloplegic such as cyclopentolate hydrochloride 1% helps reduce pain by relaxing the ciliary body. The eye should be reexamined the next day. Minor abrasions may not require patching, antibiotics, or cycloplegia.

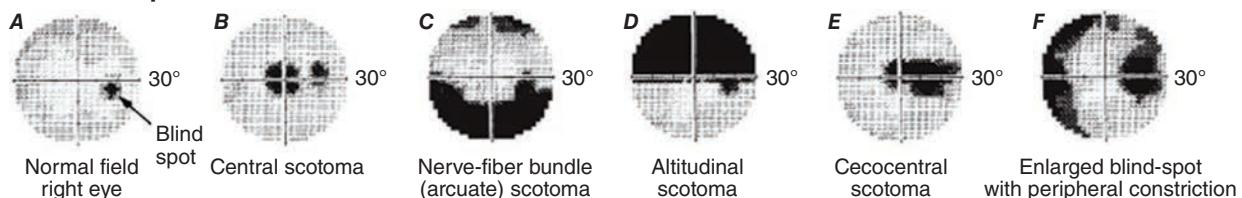
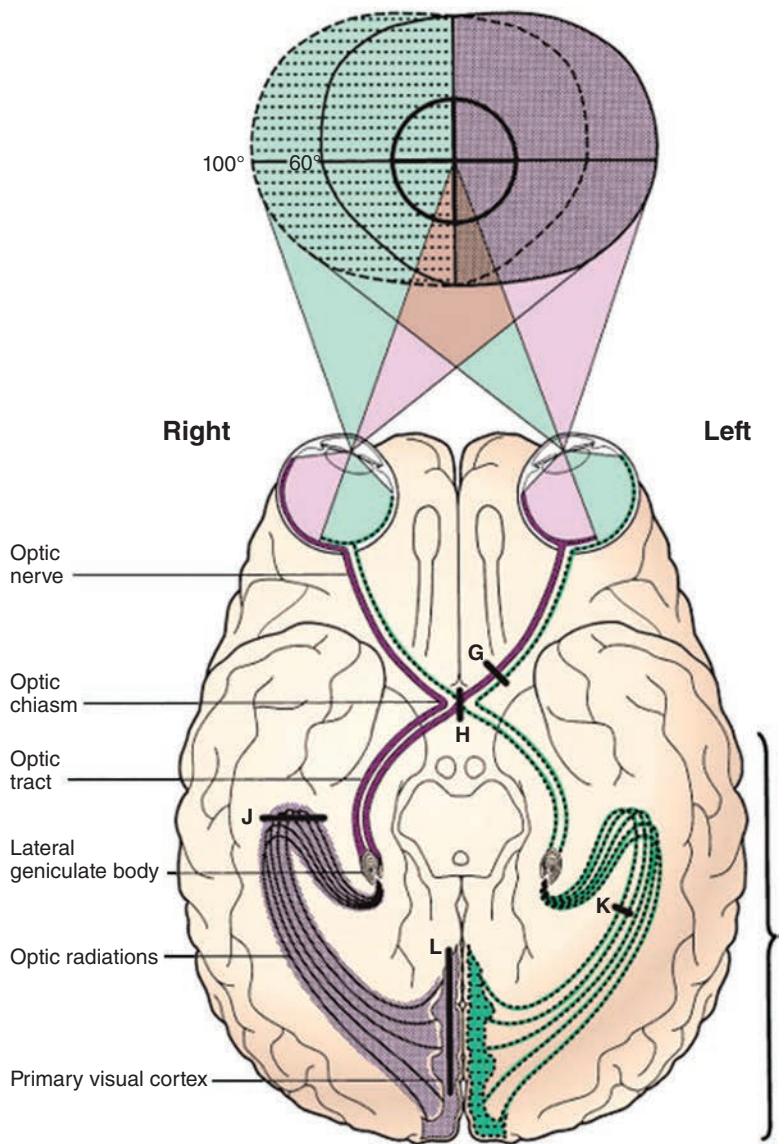
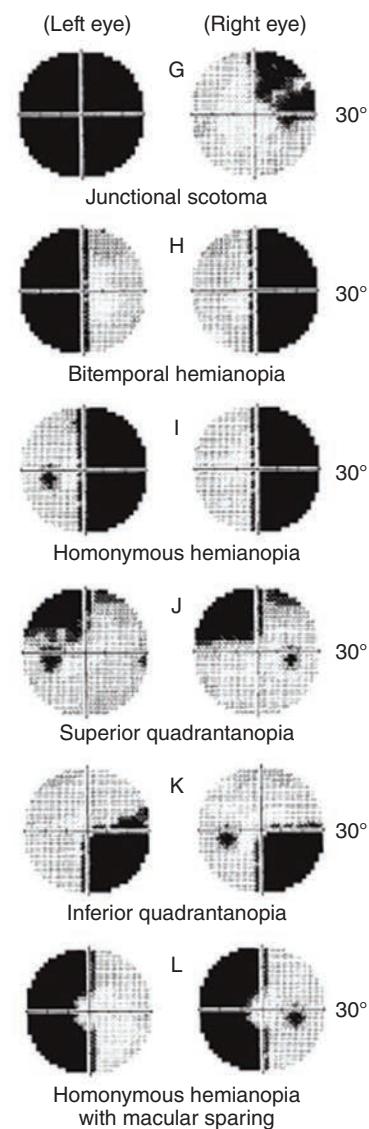
Monocular prechiasmal field defects:**Binocular chiasmal or postchiasmal field defects:**

FIGURE 32-3 Ventral view of the brain, correlating patterns of visual field loss with the sites of lesions in the visual pathway. The visual fields overlap partially, creating 120° of central binocular field flanked by a 40° monocular crescent on either side. The visual field maps in this figure were done with a computer-driven perimeter (Humphrey Instruments, Carl Zeiss, Inc.). It plots the retinal sensitivity to light in the central 30° by using a gray scale format. Areas of visual field loss are shown in black. The examples of common monocular, prechiasmal field defects are all shown for the right eye. By convention, the visual fields are always recorded with the left eye's field on the left and the right eye's field on the right, just as the patient sees the world.

Subconjunctival Hemorrhage This results from rupture of small vessels bridging the potential space between the episclera and the conjunctiva. Blood dissecting into this space can produce a spectacular red eye, but vision is not affected and the hemorrhage resolves without treatment. Subconjunctival hemorrhage is usually spontaneous but can result from blunt trauma, eye rubbing, or vigorous coughing. Occasionally, it is a clue to an underlying bleeding disorder.

Pinguecula Pinguecula is a small, raised conjunctival nodule, usually at the nasal limbus. In adults, such lesions are extremely common

and have little significance unless they become inflamed (pingueculitis). They are more apt to occur in workers with outdoor exposure. A *pterygium* resembles a pinguecula but has crossed the limbus to encroach on the corneal surface. Removal is justified when symptoms of irritation or blurring develop, but recurrence is common.

Blepharitis This refers to inflammation of the eyelids. The most common form occurs in association with acne rosacea or seborrheic dermatitis. The eyelid margins usually are colonized heavily by staphylococci. Upon close inspection, they appear greasy, ulcerated, and

crusted with scaling debris that clings to the lashes. Treatment consists of strict eyelid hygiene, applying warm compresses, and eyelash scrubs with baby shampoo. An external *hordeolum* (sty) is caused by staphylococcal infection of the superficial accessory glands of Zeis or Moll located in the eyelid margins. An internal hordeolum occurs after suppurative infection of the oil-secreting meibomian glands within the tarsal plate of the eyelid. Topical antibiotics such as bacitracin/polymyxin B ophthalmic ointment can be applied. Systemic antibiotics, usually tetracyclines or azithromycin, sometimes are necessary for treatment of meibomian gland inflammation (meibomitis) or chronic, severe blepharitis. A *chalazion* is a painless, chronic granulomatous inflammation of a meibomian gland that produces a pealike nodule within the eyelid. It can be incised and drained, but injection with glucocorticoids is equally effective. Basal cell, squamous cell, or meibomian gland carcinoma should be suspected with any nonhealing ulcerative lesion of the eyelids.

Dacryocystitis An inflammation of the lacrimal drainage system, dacryocystitis can produce epiphora (tearing) and ocular injection. Gentle pressure over the lacrimal sac evokes pain and reflux of mucus or pus from the tear puncta. Dacryocystitis usually occurs after obstruction of the lacrimal system. It is treated with topical and systemic antibiotics, followed by probing, silicone stent intubation, or surgery to reestablish patency. *Entropion* (inversion of the eyelid) or *ectropion* (sagging or eversion of the eyelid) can also lead to epiphora and ocular irritation.

Conjunctivitis Conjunctivitis is the most common cause of a red, irritated eye. Pain is minimal, and visual acuity is reduced only slightly. The most common viral etiology is adenovirus infection. It causes a watery discharge, a mild foreign-body sensation, and photophobia. Bacterial infection tends to produce a more mucopurulent exudate. Mild cases of infectious conjunctivitis usually are treated empirically with broad-spectrum topical ocular antibiotics such as sulfacetamide 10%, polymyxin-bacitracin, or a trimethoprim-polymyxin combination. Smears and cultures usually are reserved for severe, resistant, or recurrent cases of conjunctivitis. To prevent contagion, patients should be admonished to wash their hands frequently, not to touch their eyes, and to avoid direct contact with others.

Allergic Conjunctivitis This condition is extremely common and often is mistaken for infectious conjunctivitis. Itching, redness, and epiphora are typical. The palpebral conjunctiva may become hypertrophic with giant excrescences called cobblestone papillae. Irritation from contact lenses or any chronic foreign body also can induce formation of cobblestone papillae. *Atopic conjunctivitis* occurs in subjects with atopic dermatitis or asthma. Symptoms caused by allergic conjunctivitis can be alleviated with cold compresses, topical vasoconstrictors, antihistamines (olopatadine), and mast cell stabilizers (cromolyn). Topical glucocorticoid solutions provide dramatic relief of immune-mediated forms of conjunctivitis, but their long-term use is ill advised because of the complications of glaucoma, cataract, and secondary infection. Topical nonsteroidal anti-inflammatory drugs (NSAIDs; ketorolac) are better alternatives.

Keratoconjunctivitis Sicca Also known as dry eye, this produces a burning foreign-body sensation, injection, and photophobia. In mild cases, the eye appears surprisingly normal, but tear production measured by wetting of a filter paper (Schirmer strip) is deficient. A variety of systemic drugs, including antihistaminic, anticholinergic, and psychotropic medications, result in dry eye by reducing lacrimal secretion. Disorders that involve the lacrimal gland directly, such as sarcoidosis and Sjögren's syndrome, also cause dry eye. Patients may develop dry eye after radiation therapy if the treatment field includes the orbits. Problems with ocular drying are also common after lesions affecting cranial nerve V or VII. Corneal anesthesia is particularly dangerous, because the absence of a normal blink reflex exposes the cornea to injury without pain to warn the patient. Dry eye is managed by frequent and liberal application of artificial tears and ocular lubricants. In severe cases, the tear puncta can be plugged or cauterized to reduce lacrimal outflow.

Keratitis Keratitis is a threat to vision because of the risk of corneal clouding, scarring, and perforation. Worldwide, the two leading causes

of blindness from keratitis are trachoma from chlamydial infection and vitamin A deficiency related to malnutrition. In the United States, contact lenses play a major role in corneal infection and ulceration. They should not be worn by anyone with an active eye infection. In evaluating the cornea, it is important to differentiate between a superficial infection (*keratoconjunctivitis*) and a deeper, more serious ulcerative process. The latter is accompanied by greater visual loss, pain, photophobia, redness, and discharge. Slit-lamp examination shows disruption of the corneal epithelium, a cloudy infiltrate or abscess in the stroma, and an inflammatory cellular reaction in the anterior chamber. In severe cases, pus settles at the bottom of the anterior chamber, giving rise to a hypopyon. Immediate empirical antibiotic therapy should be initiated after corneal scrapings are obtained for Gram's stain, Giemsa stain, potassium hydroxide (KOH) prep, and cultures. Fortified topical antibiotics are most effective, supplemented with subconjunctival antibiotics as required. A fungal etiology should always be considered in a patient with keratitis. Fungal infection is common in warm humid climates, especially after penetration of the cornea by plant or vegetable material. Acanthamoeba keratitis is associated with improper disinfection of contact lenses.

Herpes Simplex The *herpesviruses* are a major cause of blindness from keratitis. Most adults in the United States have serum antibodies to herpes simplex, indicating prior viral infection (Chap. 192). Primary ocular infection generally is caused by herpes simplex type 1 rather than type 2. It manifests as a unilateral follicular blepharoconjunctivitis that is easily confused with adenoviral conjunctivitis, unless telltale vesicles are present on the eyelids or conjunctiva. Recurrent ocular infection arises from reactivation of latent herpesvirus. A dendritic pattern of corneal epithelial ulceration revealed by fluorescein staining is pathognomonic for herpes infection but often not present. Involvement of both eyes is extremely rare. Corneal stromal inflammation produces edema, vascularization, and iridocyclitis. Herpes keratitis is treated with cycloplegia and either a topical antiviral (trifluridine, ganciclovir) or an oral antiviral (acyclovir, valacyclovir) agent. Topical glucocorticoids are effective in mitigating corneal scarring but generally are reserved for cases involving stromal damage. Risks include corneal melting, perforation, prolonged infection, and glaucoma.

Herpes Zoster Herpes zoster from reactivation of latent varicella (chickenpox) virus causes a dermatomal pattern of painful vesicular dermatitis (Chap. 193). Ocular symptoms can occur after zoster eruption in any branch of the trigeminal nerve but are particularly common when vesicles form on the nose, reflecting nasociliary (V1) nerve involvement (Hutchinson's sign). Herpes zoster ophthalmicus produces corneal dendrites, which can be difficult to distinguish from those seen in herpes simplex. Stromal keratitis, anterior uveitis, raised intraocular pressure, ocular motor nerve palsies, acute retinal necrosis, and postherpetic scarring and neuralgia are other common sequelae. Herpes zoster ophthalmicus is treated with antiviral agents and cycloplegics. In severe cases, glucocorticoids may be added to prevent permanent visual loss from corneal scarring. Shingles should be prevented by vaccination of all healthy adults aged 50 years and older.

Episcleritis This is an inflammation of the episclera, a thin layer of connective tissue between the conjunctiva and the sclera. Episcleritis resembles conjunctivitis, but it is a more localized process and discharge is absent. Most cases of episcleritis are idiopathic, but some occur in the setting of an autoimmune disease. *Scleritis* refers to a deeper, more severe inflammatory process that frequently is associated with a connective tissue disease such as rheumatoid arthritis, lupus erythematosus, polyarteritis nodosa, granulomatosis with polyangiitis, or relapsing polychondritis. The inflammation and thickening of the sclera can be diffuse or nodular. In anterior forms of scleritis, the globe assumes a violet hue and the patient complains of severe ocular tenderness and pain. With posterior scleritis, the pain and redness may be less marked, but there is often proptosis, choroidal effusion, reduced motility, and visual loss. Episcleritis and scleritis should be treated with NSAIDs. If these agents fail, topical or even systemic glucocorticoid therapy may be necessary, especially if an underlying autoimmune process is active.

Anterior Uveitis Involving the anterior structures of the eye, uveitis was previously called *iritis* or *iridocyclitis*. The diagnosis requires slit-lamp examination to identify inflammatory cells floating in the aqueous humor or deposited on the corneal endothelium (keratic precipitates). Anterior uveitis develops in sarcoidosis, ankylosing spondylitis, juvenile idiopathic arthritis, inflammatory bowel disease, psoriasis, reactive arthritis, and Behçet's disease. It also is associated with herpes infections, syphilis, Lyme disease, onchocerciasis, tuberculosis, and leprosy. Although anterior uveitis can occur in conjunction with many diseases, no cause is found to explain the majority of cases. For this reason, laboratory evaluation usually is reserved for patients with recurrent or severe anterior uveitis. Treatment is aimed at reducing inflammation and scarring by judicious use of topical glucocorticoids. Dilatation of the pupil reduces pain and prevents the formation of synechiae.

Posterior Uveitis This diagnosis is made by observing inflammation of the vitreous, retina, or choroid on fundus examination. It is more likely than anterior uveitis to be associated with an identifiable systemic disease. Some patients have panuveitis, or inflammation of both the anterior and posterior segments of the eye. Posterior uveitis is a manifestation of autoimmune diseases such as sarcoidosis, Behçet's disease, Vogt-Koyanagi-Harada syndrome, and inflammatory bowel disease. It also accompanies diseases such as toxoplasmosis, onchocerciasis, cysticercosis, coccidioidomycosis, toxocariasis, and histoplasmosis; infections caused by organisms such as *Candida*, *Pneumocystis carinii*, *Cryptococcus*, *Aspergillus*, herpes, and cytomegalovirus (see Fig. 195-1); and other diseases, such as syphilis, Lyme disease, tuberculosis, cat-scratch disease, Whipple's disease, and brucellosis. In multiple sclerosis, chronic inflammatory changes can develop in the extreme periphery of the retina (pars planitis or intermediate uveitis). Glucocorticoids have been the mainstay of treatment for noninfectious uveitis. Biologic agents that target proinflammatory cytokines, such as the tumor necrosis factor alpha (TNF- α) inhibitor adalimumab, are effective at preventing vision loss in chronic uveitis.

Acute Angle-Closure Glaucoma This is an unusual but frequently misdiagnosed cause of a red, painful eye. Asian populations have a particularly high risk of angle-closure glaucoma. Susceptible eyes have a shallow anterior chamber because the eye has either a short axial length (hyperopia) or a lens enlarged by the gradual development of cataract. When the pupil becomes mid-dilated, the peripheral iris blocks aqueous outflow via the anterior chamber angle and the intraocular pressure rises abruptly, producing pain, injection, corneal edema, obscurations, and blurred vision. In some patients, ocular symptoms are overshadowed by nausea, vomiting, or headache, prompting a fruitless workup for abdominal or neurologic disease. The diagnosis is made by measuring the intraocular pressure during an acute attack or by performing gonioscopy, a procedure that allows one to observe a narrow chamber angle with a mirrored contact lens. Acute angle closure is treated with acetazolamide (PO or IV), topical beta blockers, prostaglandin analogues, α_2 -adrenergic agonists, and pilocarpine to induce miosis. If these measures fail, a laser can be used to create a hole in the peripheral iris to relieve pupillary block. Many physicians are reluctant to dilate patients routinely for fundus examination because they fear precipitating an angle-closure glaucoma. The risk is actually remote and more than outweighed by the potential benefit to patients of discovering a hidden fundus lesion visible only through a fully dilated pupil. Moreover, a single attack of angle closure after pharmacologic dilatation rarely causes any permanent damage to the eye and serves as an inadvertent provocative test to identify patients with narrow angles who would benefit from prophylactic laser iridectomy.

Endophthalmitis This results from bacterial, viral, fungal, or parasitic infection of the internal structures of the eye. It usually is acquired by hematogenous seeding from a remote site. Chronically ill, diabetic, or immunosuppressed patients, especially those with a history of indwelling IV catheters or positive blood cultures, are at greatest risk for endogenous endophthalmitis. Although most patients have ocular pain and injection, visual loss is sometimes the only symptom. Septic



FIGURE 32-4 Roth's spot, cotton-wool spot, and retinal hemorrhages in a 48-year-old liver transplant patient with candidemia from immunosuppression.

emboli from a diseased heart valve or a dental abscess that lodge in the retinal circulation can give rise to endophthalmitis. White-centered retinal hemorrhages known as Roth's spots (Fig. 32-4) are considered pathognomonic for subacute bacterial endocarditis, but they also appear in leukemia, diabetes, and many other conditions. Endophthalmitis occurs as a complication of ocular surgery, especially glaucoma filtering, occasionally months or even years after the operation. An occult penetrating foreign body or unrecognized trauma to the globe should be considered in any patient with unexplained intraocular infection or inflammation.

■ TRANSIENT OR SUDDEN VISUAL LOSS

Amaurosis Fugax This term refers to a transient ischemic attack of the retina (Chap. 427). Because neural tissue has a high rate of metabolism, interruption of blood flow to the retina for more than a few seconds results in *transient monocular blindness*, a term used interchangeably with amaurosis fugax. Patients describe a rapid fading of vision like a curtain descending, sometimes affecting only a portion of the visual field. Amaurosis fugax usually results from an embolus that becomes stuck within a retinal arteriole (Fig. 32-5). If the embolus breaks up or passes, flow is restored and vision returns quickly to normal without permanent damage. With prolonged interruption of blood flow, the inner retina suffers infarction. Ophthalmoscopy reveals zones of whitened, edematous retina following the distribution of branch retinal arterioles. Complete occlusion of the central retinal artery



FIGURE 32-5 Hollenhorst plaque lodged at the bifurcation of a retinal arteriole proves that a patient is shedding emboli from the carotid artery, great vessels, or heart.



FIGURE 32-6 Central retinal artery occlusion in a 78-year-old man reducing acuity to counting fingers in the right eye. Note the splinter hemorrhage on the optic disc and the slightly milky appearance to the macula with a cherry-red fovea.

produces arrest of blood flow and a milky retina with a cherry-red fovea (**Fig. 32-6**). Emboli are composed of cholesterol (Hollenhorst plaque), calcium, or platelet-fibrin debris. The most common source is an atherosclerotic plaque in the carotid artery or aorta, although emboli also can arise from the heart, especially in patients with diseased valves, atrial fibrillation, or wall motion abnormalities.

In rare instances, amaurosis fugax results from low central retinal artery perfusion pressure in a patient with a critical stenosis of the ipsilateral carotid artery and poor collateral flow via the circle of Willis. In this situation, amaurosis fugax develops when there is a dip in systemic blood pressure or a slight worsening of the carotid stenosis. Sometimes there is contralateral motor or sensory loss, indicating concomitant hemispheric cerebral ischemia.

Retinal arterial occlusion also occurs rarely in association with retinal migraine, lupus erythematosus, anticardiolipin antibodies, anticoagulant deficiency states (protein S, protein C, and antithrombin deficiency), Susac's syndrome, pregnancy, IV drug abuse, blood dyscrasias, dysproteinemias, and temporal arteritis.

Marked *systemic hypertension* causes sclerosis of retinal arterioles, splinter hemorrhages, focal infarcts of the nerve fiber layer (cotton-wool spots), and leakage of lipid and fluid (hard exudate) into the macula (**Fig. 32-7**). In hypertensive crisis, sudden visual loss can result from ischemia induced by vasospasm of retinal arterioles. In addition, visual loss can occur from ischemic optic disc swelling. Patients with acute hypertensive

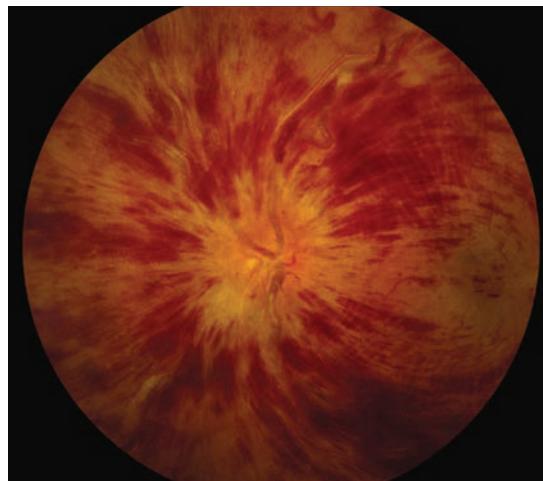


FIGURE 32-8 Central retinal vein occlusion can produce massive retinal hemorrhage ("blood and thunder"), ischemia, and vision loss.

retinopathy should be treated by lowering the blood pressure. However, the blood pressure should not be reduced precipitously, because there is a danger of optic disc infarction from sudden hypoperfusion.

Impending branch or *central retinal vein occlusion* can produce prolonged visual obscurations that resemble those described by patients with amaurosis fugax. The veins appear engorged and phlebitic, with numerous retinal hemorrhages (**Fig. 32-8**). In some patients, venous blood flow recovers spontaneously, whereas others evolve a frank obstruction with extensive retinal bleeding ("blood and thunder" appearance), infarction, and visual loss. Venous occlusion of the retina is often idiopathic, but hypertension, diabetes, and glaucoma are prominent risk factors. Polycythemia, thrombocytopenia, or other factors leading to an underlying hypercoagulable state should be corrected; aspirin treatment may be beneficial.

Anterior Ischemic Optic Neuropathy (AION) This is caused by insufficient blood flow through the posterior ciliary arteries that supply the optic disc. It produces painless monocular visual loss that is sudden in onset, followed sometimes by stuttering progression. The optic disc is edematous and usually bordered by nerve fiber layer splinter hemorrhages (**Fig. 32-9**). AION is divided into two forms: arteritic and nonarteritic. The nonarteritic form is most common. No specific cause is known, although diabetes, renal failure, and hypertension are common risk factors. Case reports have linked erectile dysfunction



FIGURE 32-7 Hypertensive retinopathy with blurred optic disc, scattered hemorrhages, cotton-wool spots (nerve fiber layer infarcts), and foveal exudate in a 62-year-old man with chronic renal failure and a systolic blood pressure of 220.



FIGURE 32-9 Anterior ischemic optic neuropathy from temporal arteritis in a 64-year-old woman with acute disc swelling, splinter hemorrhages, visual loss, and an erythrocyte sedimentation rate of 60 mm/h.

drugs to AION, but a causal association is doubtful. Evidence is strong that a crowded disc architecture and small optic cup predispose to the development of nonarteritic AION. In patients with such a “disc-at-risk,” the advent of AION in one eye increases the likelihood of the same event occurring in the other eye. No treatment is available for nonarteritic AION; glucocorticoids should not be prescribed.

About 5% of patients, especially Caucasian females aged >60, have the arteritic form of AION in conjunction with giant cell (temporal) arteritis ([Chap. 363](#)). It is urgent to recognize arteritic AION so that high doses of glucocorticoids can be instituted immediately to prevent blindness in the second eye. Tocilizumab, a monoclonal antibody against interleukin 6 receptor, is an effective alternative to glucocorticoids for sustained suppression of symptoms of giant cell arteritis. Symptoms of polymyalgia rheumatica may be present; the sedimentation rate and C-reactive protein level are usually elevated. In a patient with visual loss from suspected arteritic AION, temporal artery biopsy is mandatory to confirm the diagnosis. Administer glucocorticoids immediately, without waiting for the biopsy to be completed. The biopsy should be obtained as soon as practical, because prolonged glucocorticoid treatment can hide inflammatory changes. It is important to harvest an arterial segment at least 3 cm long and to examine a sufficient number of tissue sections. The histologic features of granulomatous inflammation are often quite subtle in temporal artery specimens. If the biopsy is declared negative by an experienced pathologist, the diagnosis of arteritic AION is highly unlikely and glucocorticoids should usually be discontinued.

Posterior Ischemic Optic Neuropathy This is an uncommon cause of acute visual loss, induced by the combination of severe anemia and hypotension. Cases have been reported after major blood loss during surgery (especially in patients undergoing cardiac or lumbar spine operations), shock, gastrointestinal bleeding, and renal dialysis. The fundus usually appears normal, although optic disc swelling develops if the process extends anteriorly far enough to reach the globe. Vision can be salvaged in some patients by immediate blood transfusion and reversal of hypotension.

Optic Neuritis This is a common inflammatory disease of the optic nerve. In the Optic Neuritis Treatment Trial (ONTT), the mean age of patients was 32 years, 77% were female, 92% had ocular pain (especially with eye movements), and 35% had optic disc swelling. In most patients, the demyelinating event was retrobulbar and the ocular fundus appeared normal on initial examination ([Fig. 32-10](#)), although optic disc pallor slowly developed over subsequent months.

Virtually all patients experience a gradual recovery of vision after a single episode of optic neuritis, even without treatment. This rule is so reliable that failure of vision to improve after a first attack of optic neuritis casts doubt on the original diagnosis. Treatment with high-dose

IV methylprednisolone (250 mg every 6 h for 3 days) followed by oral prednisone (1 mg/kg per day for 11 days) makes no difference in ultimate acuity 6 months after the attack, but the recovery of visual function occurs more rapidly. Therefore, when visual loss is severe (worse than 20/100), IV followed by PO glucocorticoids are often recommended.

For some patients, optic neuritis remains an isolated event. However, the ONTT showed that the 15-year cumulative probability of developing clinically definite multiple sclerosis after optic neuritis is 50%. A brain magnetic resonance (MR) scan is advisable in every patient with a first attack of optic neuritis. If two or more plaques are present on initial imaging, treatment should be considered to prevent the development of additional demyelinating lesions ([Chap. 444](#)).

A particularly severe optic neuritis, often involving a long segment of nerve, occurs in neuromyelitis optica (NMO); it may be bilateral and associated with myelitis. NMO can occur as a primary disorder, in the setting of systemic autoimmune disease, or rarely, as a paraneoplastic condition. Detection of circulating antibodies directed against aquaporin-4 or myelin oligodendrocyte glycoprotein (MOG) is diagnostic. Treatment for acute episodes consists of glucocorticoids followed by satralizumab, eculizumab, or inebilizumab to prevent relapse. **Neuromyelitis optica is discussed in detail in Chap. 445.**

■ LEBER'S HEREDITARY OPTIC NEUROPATHY

This disease usually affects young men, causing progressive, painless, severe central visual loss in one eye, followed weeks to years later by the same process in the other eye. Acutely, the optic disc appears mildly plethoric with surface capillary telangiectasias but no vascular leakage on fluorescein angiography. Eventually, optic atrophy ensues. Leber's optic neuropathy is caused by a point mutation at codon 11778 in the mitochondrial gene encoding nicotinamide adenine dinucleotide dehydrogenase (NADH) subunit 4. Additional mutations responsible for the disease have been identified, most in mitochondrial genes that encode proteins involved in electron transport. Mitochondrial mutations that cause Leber's neuropathy are maternally inherited by all children, but for unknown reasons, only 10% of cases occur in females. Clinical trials of gene therapy for this condition have been unsuccessful.

Toxic Optic Neuropathy This can result in acute visual loss with bilateral optic disc swelling and cecocentral scotomas. Cases have been reported from exposure to ethambutol, methyl alcohol (moonshine), ethylene glycol (antifreeze), or carbon monoxide. In toxic optic neuropathy, visual loss also can develop gradually and produce optic atrophy ([Fig. 32-11](#)) without a phase of acute optic disc edema. Many agents have been implicated in toxic optic neuropathy, but evidence supporting the association is often weak. The following is a partial list of potential offending drugs or toxins: disulfiram, ethchlorvynol, chloramphenicol,

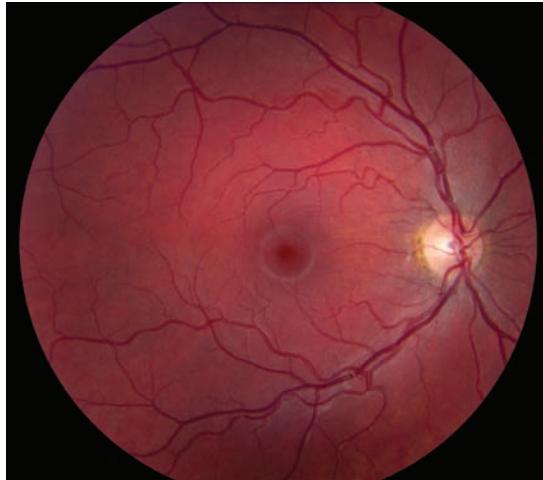


FIGURE 32-10 Retrobulbar optic neuritis is characterized by a normal fundus examination initially, hence the rubric “the doctor sees nothing, and the patient sees nothing.” Optic atrophy develops after severe or repeated attacks.



FIGURE 32-11 Optic atrophy is not a specific diagnosis but refers to the combination of optic disc pallor, arteriolar narrowing, and nerve fiber layer destruction produced by a host of eye diseases, especially optic neuropathies.



FIGURE 32-12 Papilledema means optic disc edema from raised intracranial pressure. This young woman developed acute papilledema, with hemorrhages and cotton-wool spots, as a rare side effect of treatment with tetracycline for acne.

amiodarone, monoclonal anti-CD3 antibody, ciprofloxacin, digitalis, streptomycin, lead, arsenic, thallium, D-penicillamine, isoniazid, emetine, and sulfonamides. Metallosis (chromium, cobalt, nickel) from hip implant failure is a rare cause of toxic optic neuropathy. Deficiency states induced by starvation, malabsorption, alcoholism, or gastric bypass can lead to insidious visual loss. Thiamine, vitamin B₁₂, and folate levels should be checked in any patient with unexplained bilateral central scotomas and optic pallor.

Papilledema This connotes bilateral optic disc swelling from raised intracranial pressure (Fig. 32-12). Headache is a common but not invariable accompaniment. All other forms of optic disc swelling (e.g., from optic neuritis or ischemic optic neuropathy) should be called “optic disc edema.” This convention is arbitrary but serves to avoid confusion. Often it is difficult to differentiate papilledema from other forms of optic disc edema by fundus examination alone. Transient visual obscurations are a classic symptom of papilledema. They occur in only one eye or simultaneously in both eyes. They usually last seconds but can persist longer. Obscurations follow abrupt shifts in posture or happen spontaneously. When obscurations are prolonged or spontaneous, the papilledema is more threatening. Visual acuity is not affected by papilledema unless the papilledema is severe, long-standing, or accompanied by macular edema and hemorrhage. Visual field testing shows enlarged blind spots and peripheral constriction (Fig. 32-3F). With unremitting papilledema, peripheral visual field loss progresses in an insidious fashion while the optic nerve develops atrophy. In this setting, reduction of optic disc swelling is an ominous sign of a dying nerve rather than an encouraging indication of resolving papilledema.

Evaluation of papilledema requires neuroimaging to exclude an intracranial lesion. Noninvasive MR vascular imaging may be useful in selected cases to search for a dural venous sinus thrombosis or an arteriovenous shunt. If neuroradiologic studies are negative, the subarachnoid opening pressure should be measured in the lateral decubitus position by lumbar puncture. Inaccurate pressure readings are a common pitfall. An elevated pressure, with normal cerebrospinal fluid, points by exclusion to the diagnosis of *pseudotumor cerebri* (idiopathic intracranial hypertension). Almost all patients are female, and most are obese. Treatment with a carbonic anhydrase inhibitor such as acetazolamide lowers intracranial pressure by reducing the production of cerebrospinal fluid and improves the visual fields. Weight reduction is vital; bariatric surgery should be considered in patients who cannot lose weight by diet control. If vision loss is severe or progressive, a shunt should be performed without delay to prevent blindness. Placement of a stent across the junction of the transverse and sigmoid dural sinuses, where stenosis is usually present, has emerged as a new treatment option. Optic nerve sheath fenestration is a less effective approach and



FIGURE 32-13 Optic disc drusen are calcified, mulberry-like deposits of unknown etiology within the optic disc, giving rise to “pseudopapilledema.”

does not address other neurologic symptoms. Occasionally, fulminant papilledema produces rapid onset of blindness. In such patients, emergency surgery should be performed to install a shunt.

Optic Disc Drusen These are refractile, glittering particles within the substance of the optic nerve head (Fig. 32-13). They are unrelated to drusen of the retina, which occur in age-related macular degeneration. Optic disc drusen are most common in people of northern European descent. Their diagnosis is obvious when they are visible on the surface of the optic disc. However, in many patients, they are hidden beneath the surface, producing pseudopapilledema. It is important to recognize optic disc drusen to avoid an unnecessary evaluation for papilledema. When optic disc drusen are buried, B-ultrasound is the most sensitive way to detect them. They appear hyperechoic because they contain calcium. They are also visible on computed tomography (CT) or optical coherence tomography (OCT), a technique for acquiring cross-section images of the retina. In most patients, optic disc drusen are an incidental, innocuous finding, but they can produce visual obscurations. On perimetry, they give rise to enlarged blind spots and arcuate scotomas from damage to the optic disc. With increasing age, drusen tend to become more exposed on the disc surface as optic atrophy develops. Hemorrhage, choroidal neovascular membrane, and AION are more likely to occur in patients with optic disc drusen. No treatment is available.

Vitreous Degeneration This occurs in all individuals with advancing age, leading to visual symptoms. Opacities develop in the vitreous, casting annoying shadows on the retina. As the eye moves, these distracting “floaters” move synchronously, with a slight lag caused by inertia of the vitreous gel. Vitreous traction on the retina causes mechanical stimulation, resulting in perception of flashing lights. This photopsia is brief and is confined to one eye, in contrast to the bilateral, prolonged scintillations of cortical migraine. Contraction of the vitreous can result in sudden separation from the retina, heralded by an alarming shower of floaters and photopsia. This process, known as *vitreous detachment*, is a common involutional event in the elderly. It is not harmful unless it damages the retina. A careful examination of the dilated fundus is important in any patient complaining of floaters or photopsia to search for peripheral tears or holes. If such a lesion is found, laser application can forestall a retinal detachment. Occasionally a tear ruptures a retinal blood vessel, causing vitreous hemorrhage and sudden loss of vision. On attempted ophthalmoscopy the fundus is hidden by a dark haze of blood. Ultrasound is required to examine the interior of the eye for a retinal tear or detachment. If the hemorrhage does not resolve spontaneously, the vitreous can be removed surgically. Vitreous hemorrhage also results from the fragile neovascular vessels that proliferate on the surface of the retina in diabetes, sickle cell anemia, and other ischemic ocular diseases.



FIGURE 32-14 Retinal detachment appears as an elevated sheet of retinal tissue with folds. In this patient, the fovea was spared, so acuity was normal, but an inferior detachment produced a superior scotoma.

Retinal Detachment This produces symptoms of floaters, flashing lights, and a scotoma in the peripheral visual field corresponding to the detachment (Fig. 32-14). If the detachment includes the fovea, there is an afferent pupil defect and the visual acuity is reduced. In most eyes, retinal detachment starts with a hole, flap, or tear in the peripheral retina (rhegmatogenous retinal detachment). Patients with peripheral retinal thinning (lattice degeneration) are particularly vulnerable to this process. Once a break has developed in the retina, liquefied vitreous is free to enter the subretinal space, separating the retina from the pigment epithelium. The combination of vitreous traction on the retinal surface and passage of fluid behind the retina leads inexorably to detachment. Patients with a history of myopia, trauma, or prior cataract extraction are at greatest risk for retinal detachment. The diagnosis is confirmed by ophthalmoscopic examination of the dilated eye.

Classic Migraine (See also Chap. 430) This usually occurs with a visual aura lasting about 20 min. In a typical attack, a small central disturbance in the field of vision marches toward the periphery, leaving a transient scotoma in its wake. The expanding border of migraine scotoma has a scintillating, dancing, or zigzag edge, resembling the bastions of a fortified city, hence the term *fortification spectra*. Patients' descriptions of fortification spectra vary widely and can be confused with amaurosis fugax. Migraine patterns usually last longer and are perceived in both eyes, whereas amaurosis fugax is briefer and occurs in only one eye. Migraine phenomena also remain visible in the dark or with the eyes closed. Generally, they are confined to either the right or the left visual hemifield, but sometimes, both fields are involved simultaneously. Patients often have a long history of stereotypic attacks. After the visual symptoms recede, headache develops in most patients.

Transient Ischemic Attacks Vertebrobasilar insufficiency may result in acute homonymous visual symptoms. Many patients mistakenly describe symptoms in the left or right eye when in fact the symptoms are occurring in the left or right hemifield of both eyes. Interruption of blood supply to the visual cortex causes a sudden fogging or graying of vision, occasionally with flashing lights or other positive phenomena that mimic migraine. Cortical ischemic attacks are briefer in duration than migraine, occur in older patients, and are not followed by headache. There may be associated signs of brainstem ischemia, such as diplopia, vertigo, numbness, weakness, and dysarthria.

Stroke Stroke occurs when interruption of blood supply from the posterior cerebral artery to the visual cortex is prolonged. The only finding on examination is a homonymous visual field defect that stops abruptly at the vertical meridian. Occipital lobe stroke usually is due to thrombotic occlusion of the vertebrobasilar system, embolus, or dissection. Lobar hemorrhage, tumor, abscess, and arteriovenous malformation are other common causes of hemianopic cortical visual loss.

Factitious (Functional, Nonorganic) Visual Loss This is claimed by hysterics or malingeringers. The latter account for the vast majority, seeking sympathy, special treatment, or financial gain by feigning loss of sight. The diagnosis is suspected when the history is atypical, physical findings are lacking or contradictory, inconsistencies emerge on testing, and a secondary motive can be identified. In our litigious society, the fraudulent pursuit of recompense has spawned an epidemic of factitious visual loss.

■ CHRONIC VISUAL LOSS

Cataract Cataract is a clouding of the lens sufficient to reduce vision. Most cataracts develop slowly as a result of aging, leading to gradual impairment of vision. The formation of cataract occurs more rapidly in patients with a history of uveitis, diabetes mellitus, ocular trauma, or vitrectomy. Cataracts are acquired in a variety of genetic diseases, such as myotonic dystrophy, neurofibromatosis type 2, and galactosemia. Radiation therapy and glucocorticoid treatment can induce cataract as a side effect. The cataracts associated with radiation or glucocorticoids have a typical posterior subcapsular location. Cataract can be detected by noting an impaired red reflex when viewing light reflected from the fundus with an ophthalmoscope or by examining the dilated eye with the slit lamp.

The only treatment for cataract is surgical extraction of the opacified lens. Millions of cataract operations are performed each year around the globe. The operation generally is done under local anesthesia on an outpatient basis. A plastic or silicone intraocular lens is placed within the empty lens capsule in the posterior chamber, substituting for the natural lens and leading to rapid recovery of sight. More than 95% of patients who undergo cataract extraction can expect an improvement in vision. In some patients, the lens capsule remaining in the eye after cataract extraction eventually turns cloudy, causing secondary loss of vision. A small opening, called a posterior capsulotomy, is made in the lens capsule with a laser to restore clarity.

Glaucoma Glaucoma is a slowly progressive, insidious optic neuropathy that usually is associated with chronic elevation of intraocular pressure. After cataract, it is the most common cause of blindness in the world. It is especially prevalent in people of African descent. The mechanism by which raised intraocular pressure injures the optic nerve is not understood. Axons entering the inferotemporal and superotemporal aspects of the optic disc are damaged first, producing typical nerve fiber bundle defects called arcuate scotomas. As fibers are destroyed, the neural rim of the optic disc shrinks and the physiologic cup within the optic disc enlarges (Fig. 32-15). This process is referred to as pathologic "cupping." The cup-to-disc diameter is expressed as a fraction (e.g., 0.2). The cup-to-disc ratio ranges widely in normal individuals, making it difficult to diagnose glaucoma reliably simply by

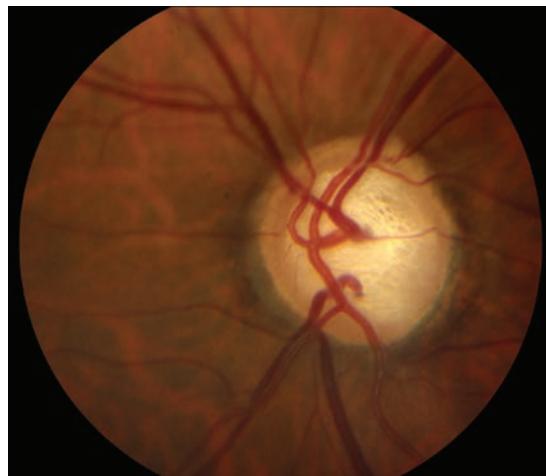


FIGURE 32-15 Glaucoma results in "cupping" as the neural rim is destroyed and the central cup becomes enlarged and excavated. The cup-to-disc ratio is about 0.8 in this patient.

observing an unusually large or deep optic cup. Careful documentation of serial examinations is helpful. In a patient with physiologic cupping, the large cup remains stable, whereas in a patient with glaucoma, it expands relentlessly over the years. Observation of progressive cupping and detection of an arcuate scotoma or a nasal step on computerized visual field testing is sufficient to establish the diagnosis of glaucoma. OCT reveals corresponding loss of fibers along the arcuate pathways in the nerve fiber layer.

The preponderance of patients with glaucoma have open anterior chamber angles. In most affected individuals, the intraocular pressure is elevated. The cause of elevated intraocular pressure is unknown, but it is associated with gene mutations in the heritable forms. Surprisingly, a third of patients with open-angle glaucoma have an intraocular pressure within the normal range of 10–20 mmHg. For this so-called normal or low-tension form of glaucoma, high myopia is a risk factor.

Chronic angle-closure glaucoma and chronic open-angle glaucoma are usually asymptomatic. Only acute angle-closure glaucoma causes a red or painful eye, from abrupt elevation of intraocular pressure. In all forms of glaucoma, foveal acuity is spared until end-stage disease is reached. For these reasons, severe and irreversible damage can occur before either the patient or the physician recognizes the diagnosis. Screening of patients for glaucoma by noting the cup-to-disc ratio on ophthalmoscopy and by measuring intraocular pressure is vital. Glaucoma is treated with topical adrenergic agonists, cholinergic agonists, beta blockers, prostaglandin analogues, and carbonic anhydrase inhibitors. Occasionally, systemic absorption of beta blocker from eyedrops can be sufficient to cause side effects of bradycardia, hypotension, heart block, bronchospasm, or depression. Laser treatment of the trabecular meshwork in the anterior chamber angle improves aqueous outflow from the eye. If medical or laser treatments fail to halt optic nerve damage from glaucoma, a filter must be constructed surgically (trabeculectomy) or a drainage device placed to release aqueous from the eye in a controlled fashion.

Macular Degeneration This is a major cause of gradual, painless, bilateral central visual loss in the elderly. It occurs in a nonexudative (dry) form and an exudative (wet) form. Inflammation may be important in both forms of macular degeneration; susceptibility is associated with variants in the gene for complement factor H, an inhibitor of the alternative complement pathway. The nonexudative process begins with the accumulation of extracellular deposits called drusen underneath the retinal pigment epithelium. On ophthalmoscopy, they are pleomorphic but generally appear as small discrete yellow lesions clustered in the macula (**Fig. 32-16**). With time, they become larger, more numerous, and confluent. The retinal pigment epithelium becomes focally detached and atrophic, causing visual loss by interfering with photoreceptor function. Treatment with vitamins C and E, beta-carotene, and zinc may retard dry macular degeneration.

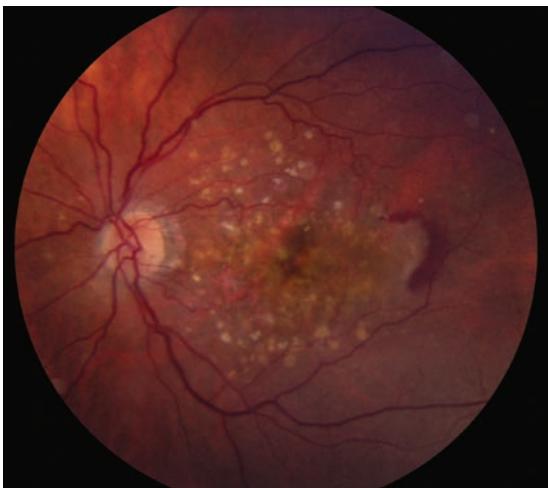


FIGURE 32-16 Age-related macular degeneration consisting of scattered yellow drusen in the macula (dry form) and a crescent of fresh hemorrhage temporal to the fovea from a subretinal neovascular membrane (wet form).

Exudative macular degeneration, which develops in only a minority of patients, occurs when neovascular vessels from the choroid grow through defects in Bruch's membrane and proliferate underneath the retinal pigment epithelium or the retina. Leakage from these vessels produces elevation of the retina, with distortion (metamorphopsia) and blurring of vision. Although the onset of these symptoms is usually gradual, bleeding from a subretinal choroidal neovascular membrane sometimes causes acute visual loss. Neovascular membranes can be difficult to see on fundus examination because they are located beneath the retina. Fluorescein angiography and OCT are extremely useful for their detection. Major or repeated hemorrhage under the retina from neovascular membranes results in fibrosis, development of a round (disciform) macular scar, and permanent loss of central vision.

A major therapeutic advance has occurred with the discovery that exudative macular degeneration can be treated with intraocular injection of antagonists to vascular endothelial growth factor. Bevacizumab, ranibizumab, afibercept, or brolucizumab is administered by direct injection into the vitreous cavity, beginning on a monthly basis. These antibodies cause the regression of neovascular membranes by blocking the action of vascular endothelial growth factor, thereby improving visual acuity.

Central Serous Chorioretinopathy This primarily affects males between the ages of 20 and 50 years. Leakage of serous fluid from the choroid causes small, localized detachment of the retinal pigment epithelium and the neurosensory retina. These detachments produce acute or chronic symptoms of metamorphopsia and blurred vision when the macula is involved. They are difficult to visualize with a direct ophthalmoscope because the detached retina is transparent and only slightly elevated. OCT shows fluid beneath the retina, and fluorescein angiography shows dye streaming into the subretinal space. The cause of central serous chorioretinopathy is unknown. Symptoms may resolve spontaneously if the retina reattaches, but recurrent detachment is common. Laser photocoagulation has benefited some patients with this condition.

Diabetic Retinopathy A rare disease until 1921, when the discovery of insulin resulted in a dramatic improvement in life expectancy for patients with diabetes mellitus, diabetic retinopathy is now a leading cause of blindness in the United States. The retinopathy takes years to develop but eventually appears in nearly all cases. Regular surveillance of the dilated fundus is crucial for any patient with diabetes. In advanced diabetic retinopathy, the proliferation of neovascular vessels leads to blindness from vitreous hemorrhage, retinal detachment, and glaucoma (**Fig. 32-17**). These complications can be avoided in most patients by administration of panretinal laser photocoagulation at the appropriate point in the evolution of the disease. Anti-vascular



FIGURE 32-17 Proliferative diabetic retinopathy in a 25-year-old man with an 18-year history of diabetes, showing neovascular vessels emanating from the optic disc, retinal and vitreous hemorrhage, cotton-wool spots, and macular exudate. Round spots in the periphery represent recently applied panretinal photocoagulation.

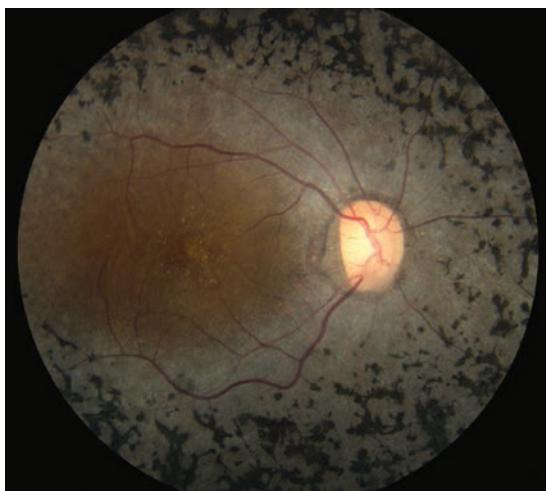


FIGURE 32-18 Retinitis pigmentosa with black clumps of pigment known as “bone spicules.” The patient had peripheral visual field loss with sparing of central (macular) vision.

endothelial growth factor antibody treatment is equally effective, but intraocular injections must be given repeatedly. **For further discussion of the manifestations and management of diabetic retinopathy, see Chaps. 403–405.**

Retinitis Pigmentosa This is a general term for a disparate group of rod-cone dystrophies characterized by progressive night blindness, visual field constriction with a ring scotoma, loss of acuity, and an abnormal electroretinogram (ERG). It occurs sporadically or in an autosomal recessive, dominant, or X-linked pattern. Irregular black deposits of clumped pigment in the peripheral retina, called *bone spicules* because of their vague resemblance to the spicules of cancellous bone, give the disease its name (Fig. 32-18). The name is actually a misnomer because retinitis pigmentosa is not an inflammatory process. Genetic testing usually identifies a mutation in the gene for rhodopsin, the rod photopigment, or in the gene for peripherin, a glycoprotein located in photoreceptor outer segments. Vitamin A (15,000 IU/d) slightly retards the deterioration of the ERG in patients with retinitis pigmentosa but has no beneficial effect on visual acuity or fields.

Leber's congenital amaurosis, a rare cone dystrophy, has been treated by replacement of the missing RPE65 protein through gene therapy, resulting in slight improvement in visual function. Some forms of retinitis pigmentosa occur in association with rare, hereditary systemic diseases (olivopontocerebellar degeneration, Bassen-Kornzweig disease, Kearns-Sayre syndrome, Refsum's disease). Chronic treatment with chloroquine, hydroxychloroquine, and phenothiazines (especially thioridazine) can produce visual loss from a toxic retinopathy that resembles retinitis pigmentosa. Patients receiving long-term treatment with hydroxychloroquine require regular eye examinations to monitor for potential development of a bull's eye maculopathy.

Epiretinal Membrane This is a fibrocellular tissue that grows across the inner surface of the retina, causing metamorphopsia and reduced visual acuity from distortion of the macula. A crinkled, cellophane-like membrane is visible on the retinal examination. Epiretinal membrane is most common in patients aged >50 years and is usually unilateral. Most cases are idiopathic, but some occur as a result of hypertensive retinopathy, diabetes, retinal detachment, or trauma. When visual acuity is reduced to the level of about 6/24 (20/80), vitrectomy and surgical peeling of the membrane to relieve macular puckering are recommended. Contraction of an epiretinal membrane sometimes gives rise to a *macular hole*. Most macular holes, however, are caused by local vitreous traction within the fovea. Vitrectomy can improve acuity in selected cases.

Melanoma and Other Tumors Melanoma is the most common primary tumor of the eye (Fig. 32-19). Approximately 2000 cases occur annually in the United States. It causes photopsia, an enlarging



FIGURE 32-19 Melanoma of the choroid, appearing as an elevated dark mass in the inferior fundus, with overlying hemorrhage. The black line denotes the plane of the optical coherence tomography scan (*below*) showing the subretinal tumor.

scotoma, and loss of vision. A small melanoma is often difficult to differentiate from a benign choroidal nevus. Serial examinations are required to document a malignant pattern of growth. Risk factors include light skin, hair, and eyes. Uveal origin accounts for 85% of cases. GNAQ and GNA11 mutations are common. About half metastasize, mainly to the liver. Small and medium-sized tumors may be treated with radiation therapy; enucleation is the best treatment for large tumors. *Metastatic tumors* to the eye outnumber primary tumors. Breast and lung carcinomas have a special propensity to spread to the choroid or iris. Leukemia and lymphoma also commonly invade ocular tissues. Sometimes their only sign on eye examination is cellular debris in the vitreous, which can masquerade as a chronic posterior uveitis.

In a patient with vision loss, CT or MR scanning should be considered if the cause remains unknown after careful review of the history, visual fields, and thorough examination of the eye. Optic nerve sheath meningioma is a common retrobulbar tumor. It produces the classic triad of optociliary shunt vessels, optic atrophy, and progressive visual loss. Optic disc swelling and proptosis are also frequent signs. Optic nerve glioma in young patients is usually a pilocytic astrocytoma and has a good prognosis for preservation of vision, especially in neurofibromatosis type 1 (Chap. 90). In adults, optic nerve glioma is rare and highly malignant. Chiasmal tumors (pituitary adenoma, meningioma, craniopharyngioma) produce visual loss with few objective findings except for optic disc pallor. Loss of the temporal visual field in each eye is typically described, but in fact, patients complain of vision loss in just one eye. A high degree of vigilance is necessary to avoid missing chiasmal tumors. Although symptoms progress gradually, in rare instances, the sudden expansion of a pituitary adenoma from infarction and bleeding (*pituitary apoplexy*) causes acute retrobulbar visual loss, with headache, nausea, and ocular motor nerve palsies.

■ PROPTOSIS

When the globes appear asymmetric, the clinician must first decide which eye is abnormal. Is one eye recessed within the orbit (*enophthalmos*), or is the other eye protuberant (*exophthalmos*, or *proptosis*)? A small globe or Horner's syndrome can give the appearance of enophthalmos. True enophthalmos occurs commonly after trauma, from atrophy of retrobulbar fat, or from fracture of the orbital floor. The position of the eyes within the orbits is measured by using a Hertel exophthalmometer, a handheld instrument that records the position of the anterior corneal surface relative to the lateral orbital rim. If this instrument is not available, relative eye position can be judged by bending the patient's head forward and looking down upon the orbits.

A proptosis of only 2 mm in one eye is detectable from this perspective. The development of proptosis implies a space-occupying lesion in the orbit and usually warrants CT or MR imaging.

Graves' Ophthalmopathy This is the leading cause of proptosis in adults (Chap. 382). The proptosis is often asymmetric and can even appear to be unilateral. Orbital inflammation and engorgement of the extraocular muscles, particularly the medial rectus and the inferior rectus, account for the protrusion of the globe. Corneal exposure, lid retraction, lid lag on downgaze, conjunctival injection, restriction of gaze, diplopia, and visual loss from optic nerve compression are cardinal symptoms. Graves' eye disease is a clinical diagnosis, but laboratory testing can be useful. The serum level of thyroid-stimulating immunoglobulins is often elevated. Orbital imaging usually reveals enlarged extraocular eye muscles, but not always. Topical lubricants, taping the eyelids closed at night, and moisture chambers are helpful to limit exposure of ocular tissues. Graves' ophthalmopathy can be treated with oral prednisone (60 mg/d) for 1 month, followed by a taper over several months, but worsening of symptoms upon glucocorticoid withdrawal is common. Infusions of teprotumumab, an inhibitor of the insulin-like growth factor I receptor, reduce proptosis and diplopia. Radiation therapy is not effective. Orbital decompression should be performed for severe, symptomatic exophthalmos or if visual function is reduced by optic nerve compression. In patients with diplopia, prisms or eye muscle surgery can be used to restore ocular alignment in primary gaze.

Orbital Pseudotumor This is an idiopathic, inflammatory orbital syndrome that is distinguished from Graves' ophthalmopathy by the prominent complaint of pain. Other symptoms include diplopia, ptosis, proptosis, and orbital congestion. Evaluation for sarcoidosis, granulomatosis with polyangiitis, and other types of orbital vasculitis or collagen-vascular disease is negative. Imaging often shows swollen eye muscles (orbital myositis) with enlarged tendons. By contrast, in Graves' ophthalmopathy, the tendons of the eye muscles usually are spared. The Tolosa-Hunt syndrome (Chap. 441) may be regarded as an extension of orbital pseudotumor through the superior orbital fissure into the cavernous sinus. The diagnosis of orbital pseudotumor is difficult. Biopsy of the orbit frequently yields nonspecific evidence of fat infiltration by lymphocytes, plasma cells, and eosinophils. A dramatic response to a therapeutic trial of systemic glucocorticoids indirectly provides the best confirmation of the diagnosis.

Orbital Cellulitis This causes pain, lid erythema, proptosis, conjunctival chemosis, restricted motility, decreased acuity, afferent pupillary defect, fever, and leukocytosis. It often arises from the paranasal sinuses, especially by contiguous spread of infection from the ethmoid sinus through the lamina papyracea of the medial orbit. A history of recent upper respiratory tract infection, chronic sinusitis, thick mucus secretions, or dental disease is significant in any patient with suspected orbital cellulitis. Blood cultures should be obtained, but they are usually negative. Most patients respond to empirical therapy with broad-spectrum IV antibiotics. Occasionally, orbital cellulitis follows an overwhelming course, with massive proptosis, blindness, septic cavernous sinus thrombosis, and meningitis. To avert this disaster, orbital cellulitis should be managed aggressively in the early stages, with immediate imaging of the orbits and antibiotic therapy that includes coverage of methicillin-resistant *Staphylococcus aureus* (MRSA). Prompt surgical drainage of an orbital abscess or paranasal sinusitis is indicated if optic nerve function deteriorates despite antibiotics.

Tumors Tumors of the orbit cause painless, progressive proptosis. The most common primary tumors are cavernous hemangioma, lymphangioma, neurofibroma, schwannoma, dermoid cyst, adenoid cystic carcinoma, optic nerve glioma, optic nerve meningioma, and benign mixed tumor of the lacrimal gland. Metastatic tumor to the orbit occurs frequently in breast carcinoma, lung carcinoma, and lymphoma. Diagnosis by fine-needle aspiration followed by urgent radiation therapy sometimes can preserve vision.

Carotid Cavernous Fistulas With anterior drainage through the orbit, these fistulas produce proptosis, diplopia, glaucoma, and

corkscrew, arterialized conjunctival vessels. Direct fistulas usually result from trauma. They are easily diagnosed because of the prominent signs produced by high-flow, high-pressure shunting. Indirect fistulas, or dural arteriovenous malformations, are more likely to occur spontaneously, especially in older women. The signs are more subtle, and the diagnosis frequently is missed. The combination of slight proptosis, diplopia, enlarged muscles, and an injected eye often is mistaken for thyroid ophthalmopathy. A bruit heard upon auscultation of the head or reported by the patient is a valuable diagnostic clue. Imaging shows an enlarged superior ophthalmic vein in the orbits. Carotid cavernous shunts can be eliminated by intravascular embolization.

■ PTOSIS

Blepharoptosis This is an abnormal drooping of the eyelid. Unilateral or bilateral ptosis can be congenital, from dysgenesis of the levator palpebrae superioris, or from abnormal insertion of its aponeurosis into the eyelid. Acquired ptosis can develop so gradually that the patient is unaware of the problem. Inspection of old photographs is helpful in dating the onset. A history of prior trauma, eye surgery, contact lens use, diplopia, systemic symptoms (e.g., dysphagia or peripheral muscle weakness), or a family history of ptosis should be sought. Fluctuating ptosis that worsens late in the day is typical of myasthenia gravis. Ptosis evaluation should focus on evidence for proptosis, eyelid masses or deformities, inflammation, pupil inequality, or limitation of motility. The width of the palpebral fissures is measured in primary gaze to determine the degree of ptosis. The ptosis will be underestimated if the patient compensates by lifting the brow with the frontalis muscle.

Mechanical Ptosis This occurs in many elderly patients from stretching and redundancy of eyelid skin and subcutaneous fat (dermatochalasis). The extra weight of these sagging tissues causes the lid to droop. Enlargement or deformation of the eyelid from infection, tumor, trauma, or inflammation also results in ptosis on a purely mechanical basis.

Aponeurotic Ptosis This is an acquired dehiscence or stretching of the aponeurotic tendon, which connects the levator muscle to the tarsal plate of the eyelid. It occurs commonly in older patients, presumably from loss of connective tissue elasticity. Aponeurotic ptosis is also a common sequela of eyelid swelling from infection or blunt trauma to the orbit, cataract surgery, or contact lens use.

Myogenic Ptosis The causes of *myogenic ptosis* include myasthenia gravis (Chap. 448) and a number of rare myopathies that manifest with ptosis. The term *chronic progressive external ophthalmoplegia* refers to a spectrum of systemic diseases caused by mutations of mitochondrial DNA. As the name implies, the most prominent findings are symmetric, slowly progressive ptosis and limitation of eye movements. In general, diplopia is a late symptom because all eye movements are reduced equally. In the *Kearns-Sayre* variant, retinal pigmentary changes and abnormalities of cardiac conduction develop. Peripheral muscle biopsy shows characteristic "ragged-red fibers." *Oculopharyngeal dystrophy* is a distinct autosomal dominant disease with onset in middle age, characterized by ptosis, limited eye movements, and trouble swallowing. *Myotonic dystrophy*, another autosomal dominant disorder, causes ptosis, ophthalmoparesis, cataract, and pigmentary retinopathy. Patients have muscle wasting, myotonia, frontal balding, and cardiac abnormalities.

Neurogenic Ptosis This results from a lesion affecting the innervation to either of the two muscles that open the eyelid: Müller's muscle or the levator palpebrae superioris. Examination of the pupil helps distinguish between these two possibilities. In Horner's syndrome, the eye with ptosis has a smaller pupil and the eye movements are full. In an oculomotor nerve palsy, the eye with the ptosis has a larger or a normal pupil. If the pupil is normal but there is limitation of adduction, elevation, and depression, a pupil-sparing oculomotor nerve palsy is likely (see next section). Rarely, a lesion affecting the small, central subnucleus of the oculomotor complex will cause bilateral ptosis with normal eye movements and pupils.

■ DOUBLE VISION (DIPLOPIA)

The first point to clarify is whether diplopia persists in either eye after the opposite eye is covered. If it does, the diagnosis is monocular diplopia. The cause is usually intrinsic to the eye and therefore has no dire implications for the patient. Corneal aberrations (e.g., keratoconus, pterygium), uncorrected refractive error, cataract, or foveal traction may give rise to monocular diplopia. Occasionally, it is a symptom of malingering or psychiatric disease. Diplopia alleviated by covering one eye is binocular diplopia and is caused by disruption of ocular alignment. Inquiry should be made into the nature of the double vision (purely side-by-side versus partial vertical displacement of images), mode of onset, duration, intermittency, diurnal variation, and associated neurologic or systemic symptoms. If the patient has diplopia while being examined, motility testing should reveal a deficiency corresponding to the patient's symptoms. However, subtle limitation of ocular excursions is often difficult to detect. For example, a patient with a slight left abducens nerve paresis may appear to have full eye movements despite a complaint of horizontal diplopia upon looking to the left. In this situation, the cover test provides a more sensitive method for demonstrating the ocular misalignment. It should be conducted in primary gaze and then with the head turned and tilted in each direction while the patient fixates a central, distant target. In the above example, a cover test with the head turned to the right bringing the eyes into left gaze will maximize the fixation shift evoked by the cover test.

Occasionally, a cover test performed in an asymptomatic patient during a routine examination will reveal an ocular deviation. If the eye movements are full and the ocular misalignment is equal in all directions of gaze (comitant deviation), the diagnosis is strabismus. In this condition, which affects about 1% of the population, fusion is disrupted in infancy or early childhood. To avoid diplopia, retinal input from the nonfixating eye may be partially suppressed. In some children, this leads to impaired vision (amblyopia, or "lazy" eye) in the deviated eye.

Binocular diplopia results from a wide range of processes: infectious, neoplastic, metabolic, degenerative, inflammatory, and vascular. One must decide whether the diplopia is neurogenic in origin or is due to restriction of globe rotation by local disease in the orbit. Orbital pseudotumor, myositis, infection, tumor, thyroid disease, and muscle entrapment (e.g., from a blowout fracture) cause restrictive diplopia. The diagnosis of restriction is usually made by recognizing other associated signs and symptoms of local orbital disease. Dedicated, high-resolution orbital imaging is helpful when the cause of diplopia is not evident.

Myasthenia Gravis (See also Chap. 448) This is a major cause of painless diplopia. The diplopia is often intermittent, variable, and not confined to any single ocular motor nerve distribution. The pupils are always normal. Serial observation of a fatigable ptosis, often accompanied by diplopia from fluctuating ocular misalignment, establishes the diagnosis. Many patients have a purely ocular form of the disease, with no evidence of systemic muscular weakness. Classically, the diagnosis was confirmed by an IV edrophonium injection, which produces a transient reversal of eyelid or eye muscle weakness, but this drug is discontinued in the United States. Blood tests for antibodies against the acetylcholine receptor or the MuSK protein are frequently negative in the purely ocular form of myasthenia gravis. *Botulism* from food or wound poisoning can mimic ocular myasthenia.

If restrictive orbital disease and myasthenia gravis are excluded, a lesion of a cranial nerve supplying innervation to the extraocular muscles is the most likely cause of binocular diplopia.

Oculomotor Nerve The third cranial nerve innervates the medial, inferior, and superior recti; inferior oblique; levator palpebrae superioris; and the iris sphincter. Total palsy of the oculomotor nerve causes ptosis, a dilated pupil, and leaves the eye "down and out" because of the unopposed action of the lateral rectus and superior oblique. This combination of findings is obvious. More challenging is the diagnosis of early or partial oculomotor nerve palsy. In this setting, any combination of ptosis, pupil dilation, and weakness of the eye muscles

supplied by the oculomotor nerve may be encountered. Frequent serial examinations during the rapidly evolving phase of the palsy help ensure that the diagnosis is not missed. The advent of an oculomotor nerve palsy with a pupil involvement, especially when accompanied by pain, suggests a compressive lesion, such as a tumor or circle of Willis aneurysm. Urgent neuroimaging should be obtained, along with a CT or MR angiogram. The resolution of these noninvasive techniques has advanced to the point that catheter angiography is rarely necessary to exclude an aneurysm.

A lesion of the oculomotor nucleus in the rostral midbrain produces signs that differ from those caused by a lesion of the nerve itself. There is bilateral ptosis because the levator muscle is innervated by a single central subnucleus. There is also weakness of the contralateral superior rectus, because it is supplied by the oculomotor nucleus on the other side. Occasionally both superior recti are weak. Isolated nuclear oculomotor palsy is rare. Usually, neurologic examination reveals additional signs that suggest brainstem damage from infarction, hemorrhage, tumor, or infection.

Injury to structures surrounding fascicles of the oculomotor nerve descending through the midbrain has given rise to a number of classic eponymic designations. In *Nothnagel's syndrome*, injury to the superior cerebellar peduncle causes ipsilateral oculomotor palsy and contralateral cerebellar ataxia. In *Benedikt's syndrome*, injury to the red nucleus results in ipsilateral oculomotor palsy and contralateral tremor, chorea, and athetosis. *Claude's syndrome* incorporates features of both of these syndromes, by injury to both the red nucleus and the superior cerebellar peduncle. Finally, in *Weber's syndrome*, injury to the cerebral peduncle causes ipsilateral oculomotor palsy with contralateral hemiparesis.

In the subarachnoid space, the oculomotor nerve is vulnerable to aneurysm, meningitis, tumor, infarction, and compression. In cerebral herniation, the nerve becomes trapped between the edge of the tentorium and the uncus of the temporal lobe. Oculomotor palsy also can result from midbrain torsion and hemorrhage during herniation. In the cavernous sinus, oculomotor palsy arises from carotid aneurysm, carotid cavernous fistula, cavernous sinus thrombosis, tumor (pituitary adenoma, meningioma, metastasis), herpes zoster infection, and the Tolosa-Hunt syndrome.

The etiology of an isolated, pupil-sparing oculomotor palsy often remains an enigma even after neuroimaging and extensive laboratory testing. Most cases are thought to result from microvascular infarction of the nerve somewhere along its course from the brainstem to the orbit. Usually, the patient complains of pain. Diabetes, hypertension, and vascular disease are major risk factors. Spontaneous recovery over a period of months is the rule. If this fails to occur or if new findings develop, the diagnosis of microvascular oculomotor nerve palsy should be reconsidered. Aberrant regeneration is common when the oculomotor nerve is injured by trauma or compression (tumor, aneurysm). Miswiring of sprouting fibers to the levator muscle and the rectus muscles results in elevation of the eyelid upon downgaze or adduction. The pupil also constricts upon attempted adduction, elevation, or depression of the globe. Aberrant regeneration is not seen after oculomotor palsy from microvascular infarct and hence vitiates that diagnosis.

Trochlear Nerve The fourth cranial nerve originates in the midbrain, just caudal to the oculomotor nerve complex. Fibers exit the brainstem dorsally and cross to innervate the contralateral superior oblique. The principal actions of this muscle are to depress and intort the globe. A palsy therefore results in hypertropia and exocyclotorsion. The cyclotorsion seldom is noticed by patients. Instead, they complain of vertical diplopia, especially upon reading or looking down. Vertical diplopia is exacerbated by tilting the head toward the side with the muscle palsy and alleviated by tilting it away. This "head tilt test" is a cardinal diagnostic feature. Review of old photographs will sometimes reveal a habitual head tilt, signifying a patient with a decompensated, congenital trochlear nerve palsy.

New, isolated trochlear nerve palsy results from all the causes listed above for the oculomotor nerve except aneurysm. The trochlear nerve is particularly apt to suffer injury after closed head trauma. The free edge of the tentorium impinges on the nerve during a concussive blow.

Most isolated trochlear nerve palsies are idiopathic and hence are diagnosed by exclusion as “microvascular.” Spontaneous improvement occurs over a period of months in most patients. A base-down prism (conveniently applied to the patient’s glasses as a stick-on Fresnel lens) may serve as a temporary measure to alleviate diplopia. If the palsy does not resolve, the eyes can be realigned by weakening the inferior oblique muscle.

Abducens Nerve The sixth cranial nerve innervates the lateral rectus muscle. A palsy produces horizontal diplopia, worse on gaze to the side of the lesion. A nuclear lesion has different consequences, because the abducens nucleus contains interneurons that project via the medial longitudinal fasciculus to the medial rectus subnucleus of the contralateral oculomotor complex. Therefore, an abducens nuclear lesion produces a complete lateral gaze palsy from weakness of both the ipsilateral lateral rectus and the contralateral medial rectus. *Foville’s syndrome* after dorsal pontine injury includes lateral gaze palsy, ipsilateral facial palsy, and contralateral hemiparesis incurred by damage to descending corticospinal fibers. *Millard-Gubler syndrome* from ventral pontine injury is similar except for the eye findings. There is lateral rectus weakness only, instead of gaze palsy, because the abducens fascicle is injured rather than the nucleus. Infarct, tumor, hemorrhage, vascular malformation, and multiple sclerosis are the most common etiologies of brainstem abducens palsy.

After leaving the ventral pons, the abducens nerve runs forward along the clivus to pierce the dura at the petrous apex, where it enters the cavernous sinus. Along its subarachnoid course, it is susceptible to meningitis, tumor (meningioma, chordoma, carcinomatous meningitis), subarachnoid hemorrhage, trauma, and compression by aneurysm or dolichoectatic vessels. At the petrous apex, mastoiditis can produce deafness, pain, and ipsilateral abducens palsy (*Gradenigo’s syndrome*). In the cavernous sinus, the nerve can be affected by carotid aneurysm, carotid cavernous fistula, tumor (pituitary adenoma, meningioma, nasopharyngeal carcinoma), herpes infection, and Tolosa-Hunt syndrome.

Unilateral or bilateral abducens palsy is a classic sign of raised intracranial pressure. The diagnosis can be confirmed if papilledema is observed on fundus examination. The mechanism is still debated but probably is related to rostral-caudal displacement of the brainstem. The same phenomenon accounts for abducens palsy from Chiari malformation or low intracranial pressure (e.g., after lumbar puncture, spinal anesthesia, or spontaneous dural cerebrospinal fluid leak).

Treatment of abducens palsy is aimed at prompt correction of the underlying cause. However, the cause remains obscure in many instances despite diligent evaluation. As was mentioned above for isolated trochlear or oculomotor palsy, most cases are assumed to represent microvascular infarcts because they often occur in the setting of diabetes or other vascular risk factors. Some cases may develop as a postinfectious mononeuritis (e.g., after a viral flu). Patching one eye, occluding one eyeglass lens with tape, or applying a temporary prism will provide relief of diplopia until the palsy resolves. If recovery is incomplete, eye muscle surgery nearly always can realign the eyes, at least in primary position. A patient with an abducens palsy that fails to improve should be reevaluated for an occult etiology (e.g., chordoma, carcinomatous meningitis, carotid cavernous fistula, myasthenia gravis). Skull base tumors are easily missed even on contrast-enhanced neuroimaging studies.

Multiple Ocular Motor Nerve Palsies These should not be attributed to spontaneous microvascular events affecting more than one cranial nerve at a time. This remarkable coincidence does occur, especially in diabetic patients, but the diagnosis is made only in retrospect after all other diagnostic alternatives have been exhausted. Neuroimaging should focus on the cavernous sinus, superior orbital fissure, and orbital apex, where all three ocular motor nerves are in close proximity. In a diabetic or immunocompromised host, fungal infection (*Aspergillus*, *Mucorales*, *Cryptococcus*) is a common cause of multiple nerve palsies. In a patient with systemic malignancy, carcinomatous meningitis is a likely diagnosis. Cytologic examination may be

negative despite repeated sampling of the cerebrospinal fluid. The cancer-associated Lambert-Eaton myasthenic syndrome also can produce ophthalmoplegia. Giant cell (temporal) arteritis occasionally manifests as diplopia from ischemic palsies of extraocular muscles. Fisher’s syndrome, an ocular variant of Guillain-Barré, produces ophthalmoplegia with areflexia and ataxia. Often the ataxia is mild, and the reflexes are normal. Antiganglioside antibodies (GQ1b) can be detected in about 50% of cases.

Supranuclear Disorders of Gaze These are often mistaken for multiple ocular motor nerve palsies. For example, Wernicke’s encephalopathy can produce nystagmus and a partial deficit of horizontal and vertical gaze that mimics a combined abducens and oculomotor nerve palsy. The disorder occurs in patients who are malnourished, alcoholic, or following bariatric surgery, and can be reversed by thiamine. Infarct, hemorrhage, tumor, multiple sclerosis, encephalitis, vasculitis, and Whipple’s disease are other important causes of supranuclear gaze palsy. Disorders of vertical gaze, especially downward saccades, are an early feature of progressive supranuclear palsy. Smooth pursuit is affected later in the course of the disease. Parkinson’s disease, Huntington’s disease, and olivopontocerebellar degeneration also can affect vertical gaze.

The *frontal eye field* of the cerebral cortex is involved in generation of saccades to the contralateral side. After hemispheric stroke, the eyes usually deviate toward the lesioned side because of the unopposed action of the frontal eye field in the normal hemisphere. With time, this deficit resolves. Seizures generally have the opposite effect: the eyes deviate conjugately away from the irritative focus. *Parietal lesions* disrupt smooth pursuit of targets moving toward the side of the lesion. Bilateral parietal lesions produce *Bálint’s syndrome*, which is characterized by impaired eye-hand coordination (optic ataxia), difficulty initiating voluntary eye movements (ocular apraxia), and visuospatial disorientation (simultanagnosia).

Horizontal Gaze Descending cortical inputs mediating horizontal gaze ultimately converge at the level of the pons. Neurons in the paramedian pontine reticular formation are responsible for controlling conjugate gaze toward the same side. They project directly to the ipsilateral abducens nucleus. A lesion of either the paramedian pontine reticular formation or the abducens nucleus causes an ipsilateral conjugate gaze palsy. Lesions at either locus produce nearly identical clinical syndromes, with the following exception: vestibular stimulation (oculocephalic maneuver or caloric irrigation) will succeed in driving the eyes conjugately to the side in a patient with a lesion of the paramedian pontine reticular formation but not in a patient with a lesion of the abducens nucleus.

INTERNUCLEAR OPHTHALMOPLEGIA This results from damage to the medial longitudinal fasciculus ascending from the abducens nucleus in the pons to the oculomotor nucleus in the midbrain (hence, “internuclear”). Damage to fibers carrying the conjugate signal from abducens interneurons to the contralateral medial rectus motoneurons results in a failure of adduction on attempted lateral gaze. For example, a patient with a left internuclear ophthalmoplegia (INO) will have slowed or absent adducting movements of the left eye (Fig. 32-20). A patient with bilateral injury to the medial longitudinal fasciculus will have bilateral INO. Multiple sclerosis is the most common cause, although tumor, stroke, trauma, or any brainstem process may be responsible. *One-and-a-half syndrome* is due to a lesion of the medial longitudinal fasciculus combined with a lesion of either the abducens nucleus or the paramedian pontine reticular formation on the same side. The patient’s only horizontal eye movement is abduction of the eye on the other side.

Vertical Gaze This is controlled at the level of the midbrain. The neuronal circuits affected in disorders of vertical gaze are not fully elucidated, but lesions of the rostral interstitial nucleus of the medial longitudinal fasciculus and the interstitial nucleus of Cajal cause supranuclear paresis of upgaze, downgaze, or all vertical eye movements. Distal basilar artery ischemia is the most common etiology.

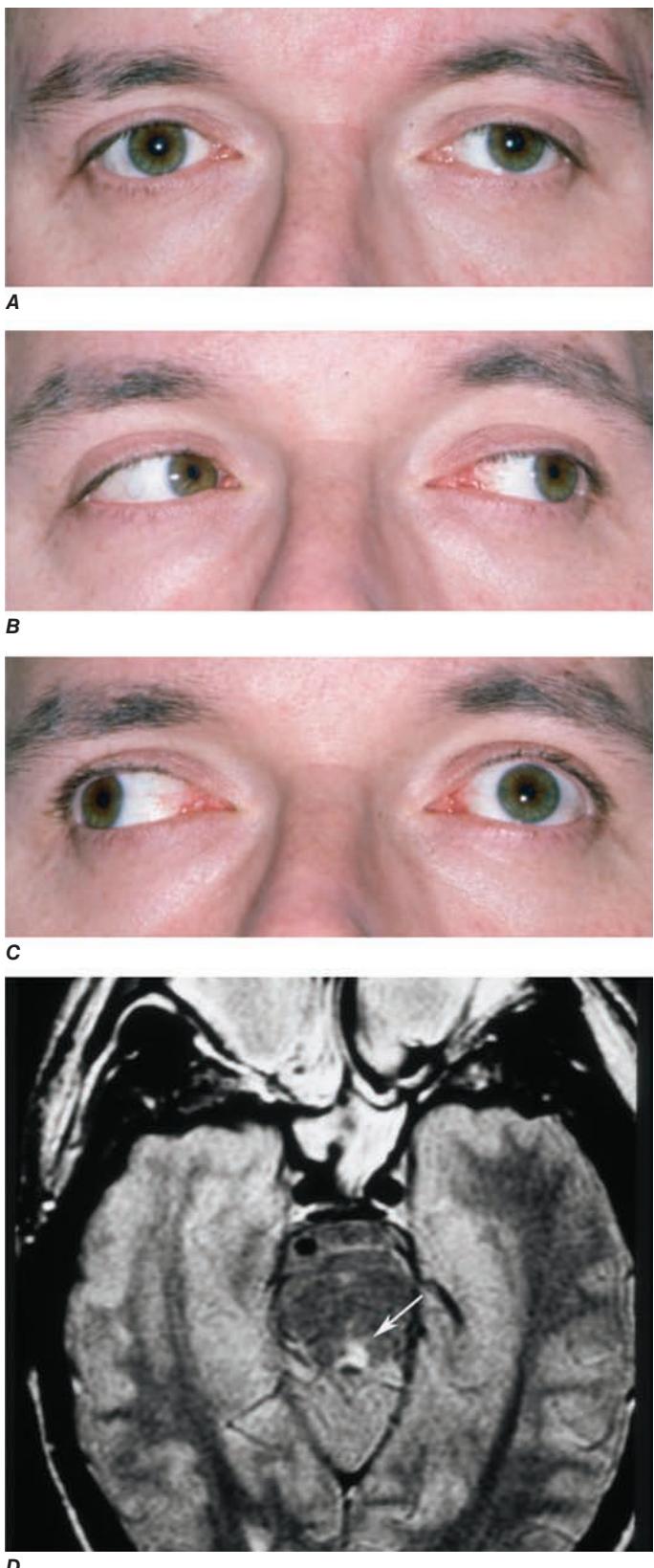


FIGURE 32-20 Left internuclear ophthalmoplegia (INO). **A.** In primary position of gaze, the eyes appear normal. **B.** Horizontal gaze to the left is intact. **C.** On attempted horizontal gaze to the right, the left eye fails to adduct. In mildly affected patients, the eye may adduct partially or more slowly than normal. Nystagmus is usually present in the abducted eye. **D.** T2-weighted axial magnetic resonance image through the pons showing a demyelinating plaque in the left medial longitudinal fasciculus (arrow).

Skew deviation refers to a vertical misalignment of the eyes, usually constant in all positions of gaze. The finding has poor localizing value because skew deviation has been reported after lesions in widespread regions of the brainstem and cerebellum.

PARINAUD'S SYNDROME Also known as dorsal midbrain syndrome, this is a distinct supranuclear vertical gaze disorder caused by damage to the posterior commissure. It is a classic sign of hydrocephalus from aqueductal stenosis. Pineal region or midbrain tumors, cysticercosis, and stroke also cause Parinaud's syndrome. Features include loss of upgaze (and sometimes downgaze), convergence-retraction nystagmus on attempted upgaze, downward ocular deviation ("setting sun" sign), lid retraction (Collier's sign), skew deviation, pseudoabducens palsy, and light-near dissociation of the pupils.

Nystagmus This is a rhythmic oscillation of the eyes, occurring physiologically from vestibular and optokinetic stimulation or pathologically in a wide variety of diseases (Chap. 22). Abnormalities of the eyes or optic nerves, present at birth or acquired in childhood, can produce a complex, searching nystagmus with irregular pendular (sinusoidal) and jerk features. Examples are albinism, Leber's congenital amaurosis, and bilateral cataract. This nystagmus is commonly referred to as *congenital sensory nystagmus*. This is a poor term because even in children with congenital lesions, the nystagmus does not appear until weeks after birth. *Congenital motor nystagmus*, which looks similar to congenital sensory nystagmus, develops in the absence of any abnormality of the sensory visual system. Visual acuity also is reduced in congenital motor nystagmus, probably by the nystagmus itself, but seldom below a level of 20/200.

JERK NYSTAGMUS This is characterized by a slow drift off the target, followed by a fast corrective saccade. By convention, the nystagmus is named after the quick phase. Jerk nystagmus can be downbeat, upbeat, horizontal (left or right), and torsional. The pattern of nystagmus may vary with gaze position. Some patients will be oblivious to their nystagmus. Others will complain of blurred vision or a subjective to-and-fro movement of the environment (oscillopsia) corresponding to the nystagmus. Fine nystagmus may be difficult to see on gross examination of the eyes. Observation of nystagmoid movements of the optic disc on ophthalmoscopy is a sensitive way to detect subtle nystagmus.

GAZE-EVOKED NYSTAGMUS This is the most common form of jerk nystagmus. When the eyes are held eccentrically in the orbits, they have a natural tendency to drift back to primary position. The subject compensates by making a corrective saccade to maintain the deviated eye position. Many normal patients have mild gaze-evoked nystagmus. Exaggerated gaze-evoked nystagmus can be induced by drugs (sedatives, anticonvulsants, alcohol); muscle paresis; myasthenia gravis; demyelinating disease; and cerebellopontine angle, brainstem, and cerebellar lesions.

VESTIBULAR NYSTAGMUS Vestibular nystagmus results from dysfunction of the labyrinth (Ménière's disease), vestibular nerve, or vestibular nucleus in the brainstem. Peripheral vestibular nystagmus often occurs in discrete attacks, with symptoms of nausea and vertigo. There may be associated tinnitus and hearing loss. Sudden shifts in head position may provoke or exacerbate symptoms.

DOWNBEAT NYSTAGMUS Downbeat nystagmus results from lesions near the craniocervical junction (Chiari malformation, basilar invagination). It also has been reported in brainstem or cerebellar stroke, lithium or anticonvulsant intoxication, alcoholism, and multiple sclerosis. Upbeat nystagmus is associated with damage to the pontine tegmentum from stroke, demyelination, or tumor.

Opsoclonus This rare, dramatic disorder of eye movements consists of bursts of consecutive saccades (saccadomania). When the saccades are confined to the horizontal plane, the term *ocular flutter* is preferred. It can result from viral encephalitis, trauma, or a paraneoplastic effect of neuroblastoma, breast carcinoma, and other malignancies. It has also been reported as a benign, transient phenomenon in otherwise healthy patients.

FURTHER READING

- ADAMIS AP et al: Building on the success of anti-vascular endothelial growth factor therapy: A vision for the next decade. *Eye* 34:1966, 2020.
- DOUGLAS RS: Teprotumumab for the treatment of active thyroid eye disease. *N Engl J Med* 382:341, 2020.
- DOWLING JE: Restoring vision to the blind. *Science* 368:827, 2020.
- GROSS JG et al: Panretinal photocoagulation vs intravitreous ranibizumab for proliferative diabetic retinopathy. *JAMA* 314:2137, 2015.
- JAFFE GJ et al: Adalimumab in patients with active noninfectious uveitis. *N Engl J Med* 375:932, 2016.
- MAEDER ML: Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat Med* 25:229, 2019.
- PIORO MH: Primary care vasculitis: Polymyalgia rheumatica and giant cell arteritis. *Prim Care* 45:305, 2018.
- STONE JH et al: Trial of tocilizumab in giant-cell arteritis. *N Engl J Med* 377:317, 2017.
- YANOFF M, DUKER J: *Ophthalmology*, 5th ed. Atlanta, Saunders, 2019.

33

Disorders of Smell and Taste

Richard L. Doty, Steven M. Bromley



All environmental chemicals necessary for life enter the body by the nose and mouth. The senses of smell (olfaction) and taste (gustation) monitor such chemicals, determine the flavor and palatability of foods and beverages, and warn of dangerous environmental conditions, including fire, air pollution, leaking natural gas, and bacteria-laden foodstuffs. These senses contribute significantly to quality of life and, when dysfunctional, can have untoward physical and psychological consequences. A longitudinal study of 1162 nondemented elderly persons found, even after controlling for confounders, that those with the lowest baseline olfactory test scores had a 45% mortality rate over a 4-year period, compared to an 18% mortality rate for those with the highest olfactory test scores. A basic understanding of these senses in health and disease is critical for the physician, because thousands of patients present to doctors' offices each year with complaints of chemosensory dysfunction. Among the more important recent developments in neurology is the discovery that decreased smell function is among the first

signs of such neurodegenerative diseases as Parkinson's disease (PD) and Alzheimer's disease (AD), signifying their "presymptomatic" phase.

ANATOMY AND PHYSIOLOGY

Olfactory System Odorous chemicals enter the front of nose during inhalation and active sniffing, as well as the back of the nose (nasopharynx) during deglutition. After reaching the highest recesses of the nasal cavity, they dissolve in the olfactory mucus and diffuse or are actively transported by specialized proteins to receptors located on the cilia of olfactory receptor cells. The cilia, dendrites, cell bodies, and proximal axonal segments of these bipolar cells are located within a unique neuroepithelium covering the cribriform plate, the superior nasal septum, superior turbinate, and sectors of the middle turbinate (Fig. 33-1). Nearly 400 types of G-protein-coupled odor receptors (GPCRs) are expressed on the cilia of the receptor cells, with only one type of GPCR being expressed on a given cell. Other receptors, including trace amine-associated receptors and members of the non-GPCR membrane-spanning 4-domain family, subfamily A (MS4A) protein family, are also present on some receptor cells. Such a plethora of receptor cell types does not exist in any other sensory system. Importantly, when damaged, the receptor cells can be replaced by stem cells near the basement membrane, although such replacement is often incomplete.

After coalescing into bundles surrounded by glia-like ensheathing cells (termed fila), the receptor cell axons pass through the cribriform plate to the olfactory bulbs, where they synapse with dendrites of other cell types within the glomeruli (Fig. 33-2). These spherical structures, which make up a distinct layer of the olfactory bulb, are a site of convergence of information, because many more fibers enter than leave them. Receptor cells that express the same type of receptor project to the same glomeruli, effectively making each glomerulus a functional unit. The major projection neurons of the olfactory system—the mitral and tufted cells—send primary dendrites into the glomeruli, connecting not only with the incoming receptor cell axons, but with dendrites of periglomerular cells. The activity of the mitral/tufted cells is modulated by the periglomerular cells, secondary dendrites from other mitral/tufted cells, and granule cells, the most numerous cells of the bulb. The latter cells, which are largely GABAergic, receive inputs from central brain structures and modulate the output of the mitral/tufted cells. Interestingly, like the olfactory receptor cells, some cells within the bulb undergo replacement. Thus, neuroblasts formed within the anterior subventricular zone of the brain migrate along the rostral migratory stream, ultimately becoming granule and periglomerular cells.

The axons of the mitral and tufted cells synapse within secondary olfactory structures, which largely compose the primary olfactory cortex (POC) (Fig. 33-3). The POC is defined as those cortical structures that receive direct projections from the olfactory bulb, most notably the piriform and entorhinal cortices. Although olfaction is unique

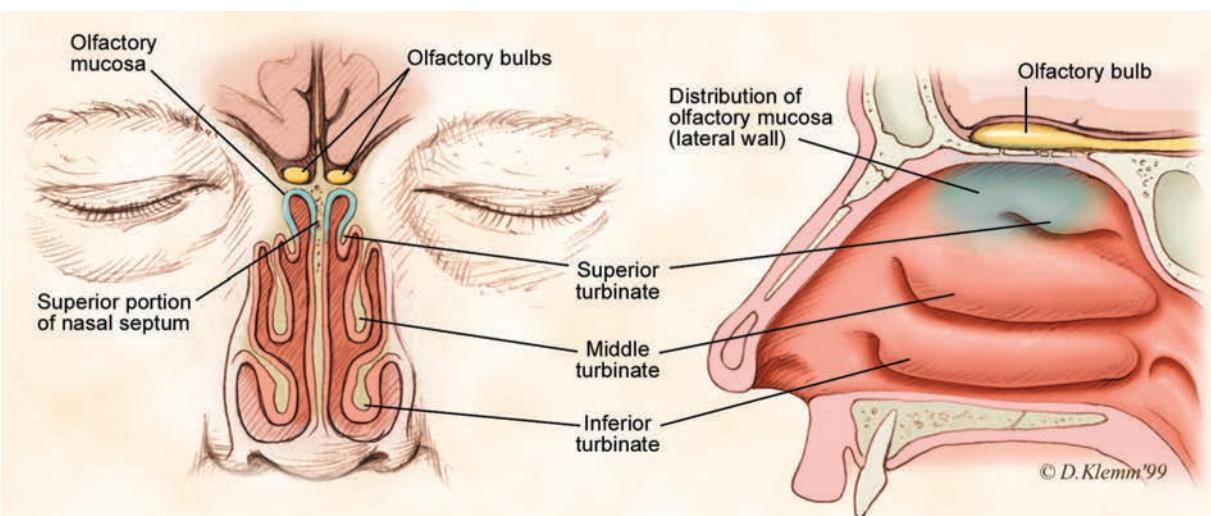


FIGURE 33-1 Anatomy of the nose, showing the distribution of olfactory receptors in the roof of the nasal cavity. (Copyright David Klemm, Faculty and Curriculum Support [FACS], Georgetown University Medical Center.)

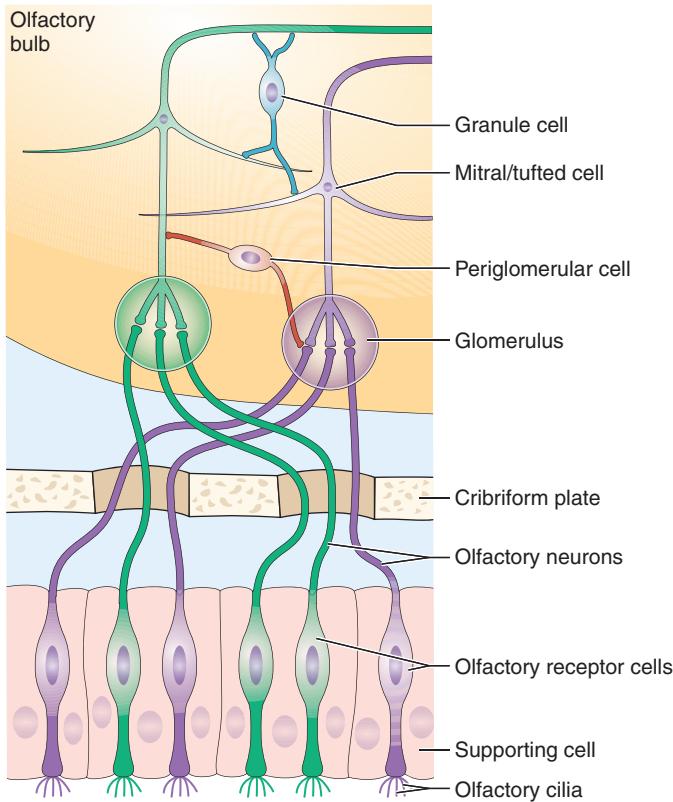


FIGURE 33-2 Schematic of the layers and wiring of the olfactory bulb. Each receptor type (red, green, blue) projects to a common glomerulus. The neural activity within each glomerulus is modulated by periglomerular cells. The activity of the primary projection cells, the mitral and tufted cells, is modulated by granule cells, periglomerular cells, and secondary dendrites from adjacent mitral and tufted cells. (Adapted from www.med.yale.edu/neurosurg/treloar/index.html)

in that its initial afferent projections bypass the thalamus, persons with damage to the thalamus can exhibit olfactory deficits, particularly ones of odor identification. Such deficits likely reflect the involvement of thalamic connections between the POC and the orbitofrontal cortex (OFC), where odor identification largely occurs. The close anatomic ties between the olfactory system and the amygdala, hippocampus,

and hypothalamus help to explain the intimate associations between odor perception and cognitive functions such as memory, motivation, arousal, autonomic activity, digestion, and sex.

Taste System Tastants are sensed by specialized receptor cells present within taste buds—small grapefruit-like segmented structures located on the lateral margins and dorsum of the tongue, roof of the mouth, pharynx, larynx, and superior esophagus (Fig. 33-4). Lingual taste buds are embedded in well-defined protuberances, termed fungiform, foliate, and circumvallate papillae. After dissolving in a liquid, tastants enter the opening of the taste bud—the taste pore—and bind to receptors on microvilli, small extensions of receptor cells within each taste bud. Such binding changes the electrical potential across the taste cell, resulting in neurotransmitter release onto the first-order taste neurons. Although humans have ~7500 taste buds, not all harbor taste-sensitive cells; some contain only one class of receptor (e.g., cells responsive only to sugars), whereas others contain cells sensitive to more than one class. The number of taste receptor cells per taste bud ranges from zero to well over 100. A small family of three GPCRs, namely T1R1, T1R2, and T1R3, mediate sweet and umami taste sensations. Bitter sensations, on the other hand, depend on T2R receptors, a family of ~30 GPCRs expressed on cells different from those that express the sweet and umami receptors. T2Rs sense a wide range of bitter substances but do not distinguish among them. Sour tastants are sensed by the PKD2L1 receptor, a member of the transient receptor potential protein (TRP) family. Perception of salty sensations, such as induced by sodium chloride, arises from the entry of Na^+ ions into the cells via specialized membrane channels, such as the amiloride-sensitive Na^+ channel.

It is now well established that both bitter and sweet taste-related receptors are also present elsewhere in the body, most notably in the alimentary and respiratory tracts. This important discovery generalizes the concept of taste-related chemoreception to areas of the body beyond the mouth and throat, with α -gustducin, the taste-specific G-protein α -subunit, expressed in so-called brush cells found specifically within the human trachea, lung, pancreas, and gallbladder. These brush cells are rich in nitric oxide (NO) synthase, known to defend against xenobiotic organisms, protect the mucosa from acid-induced lesions, and, in the case of the gastrointestinal tract, stimulate vagal and splanchnic afferent neurons. NO further acts on nearby cells, including enteroendocrine cells, absorptive or secretory epithelial cells, mucosal blood vessels, and cells of the immune system. Members of the T2R family of bitter receptors and the sweet receptors of the T1R family have been identified within the gastrointestinal tract and in enteroendocrine cell lines. In some cases, these receptors are important for metabolism, with the T1R3 receptors and gustducin playing decisive roles in the sensing and transport of dietary sugars from the intestinal lumen into absorptive enterocytes via a sodium-dependent glucose transporter and in regulation of hormone release from gut enteroendocrine cells. In other cases, these receptors may be important for airway protection, with a number of T2R bitter receptors in the motile cilia of the human airway that respond to bitter compounds by increasing their beat frequency. One specific T2R38 taste receptor is expressed in human upper respiratory epithelia and responds to acyl-monoserine lactone quorum-sensing molecules secreted by *Pseudomonas aeruginosa* and other gram-negative bacteria. Differences in T2R38 functionality, as related to TAS2R38 genotype, correlate with susceptibility to upper respiratory infections in humans.

Taste information is sent to the brain via three cranial nerves (CNs): CN VII (the *facial nerve*, which involves the intermediate nerve with its branches, the greater petrosal and chorda tympani nerves), CN IX (the *glossopharyngeal nerve*), and CN X (the *vagus nerve*) (Fig. 33-5). CN VII innervates the anterior tongue and all of the soft palate, CN IX innervates the posterior tongue, and CN X innervates the laryngeal surface of the epiglottis, larynx, and proximal portion of the esophagus. The mandibular branch of CN V (V_3) conveys somatosensory information (e.g., touch, burning, cooling, irritation) to the brain. Although not technically a gustatory nerve, CN V shares primary nerve routes with many of the gustatory nerve fibers and adds temperature, texture, pungency, and spiciness to the taste experience. The chorda tympani

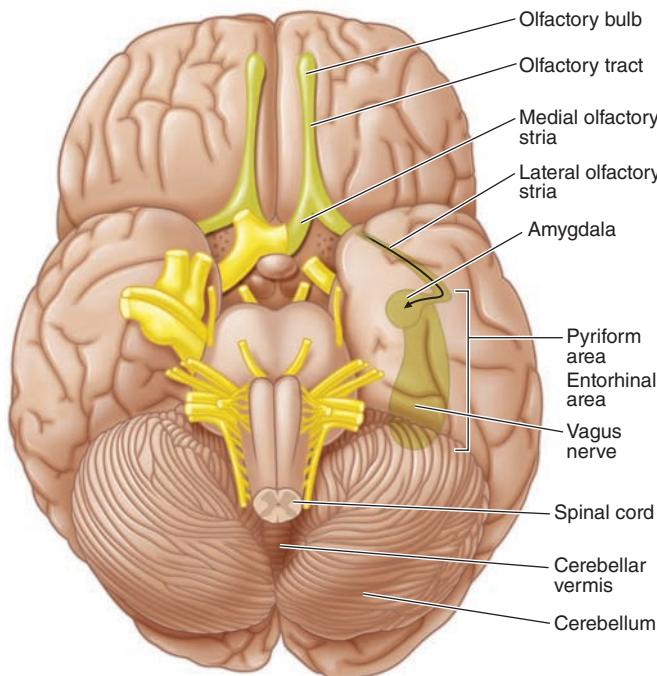


FIGURE 33-3 Anatomy of the base of the brain showing the primary olfactory cortex.

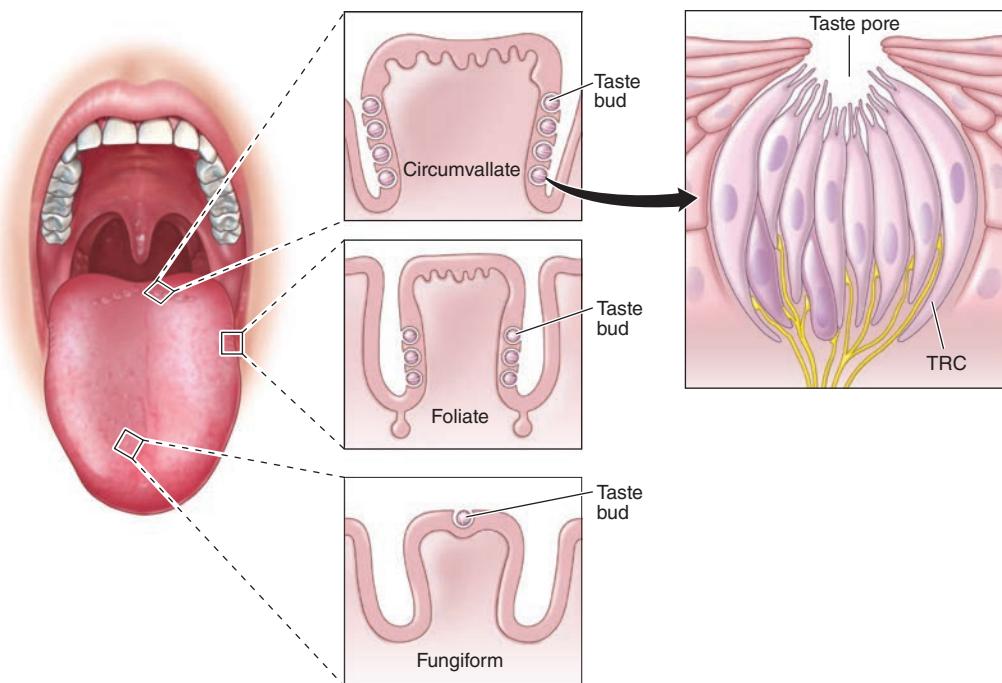


FIGURE 33-4 Schematic of the taste bud and its opening (pore), as well as the location of buds on the three major types of papillae: fungiform (anterior), foliate (lateral), and circumvallate (posterior). TRC, taste receptor cell.

nerve is famous for taking a recurrent course through the facial canal in the petrosal portion of the temporal bone, passing through the middle ear, and then exiting the skull via the petrotympanic fissure, where it joins the lingual nerve (a division of CN V) near the tongue. This nerve also carries parasympathetic fibers to the submandibular and sublingual glands, whereas the greater petrosal nerve supplies the palatine glands, thereby influencing saliva production.

The axons of the projection cells, which synapse with taste buds, enter the rostral portion of the nucleus of the solitary tract (NTS) within the medulla of the brainstem (Fig. 33-5). From the NTS, neurons then project to a division of the ventroposteromedial thalamic nucleus (VPM) via the medial lemniscus. From here, projections are made to the rostral part of the frontal operculum and adjoining insula,

a brain region considered the *primary taste cortex* (PTC). Projections from the PTC then go to the *secondary taste cortex*, namely the caudolateral OFC. This brain region is involved in the conscious recognition of taste qualities. Moreover, because it contains cells that are activated by several sensory modalities, it is likely a center for establishing “flavor.”

DISORDERS OF OLFACTION

The ability to smell is influenced, in everyday life, by such factors as age, gender, general health, nutrition, smoking, and reproductive state. Women typically outperform men on tests of olfactory function and retain normal smell function to a later age than do men.

Estimates of the prevalence of olfactory dysfunction in the general population vary; a cross-sectional analysis from the National Health and Nutrition Examination Survey (NHANES 2013–2014) found an overall prevalence of 13.5%. However, it is apparent that significant decrements in the ability to smell are present in >50% of the population between 65 and 80 years of age and in 75% of those aged ≥80 years (Fig. 33-6). Such presbyosmia helps to explain why many elderly report

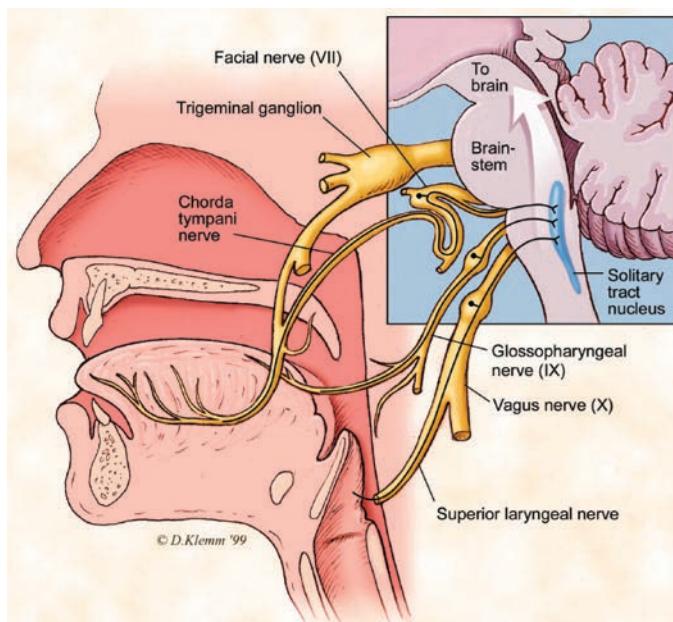


FIGURE 33-5 Schematic of the cranial nerves (CNs) that mediate taste function, including the chorda tympani nerve (CN VII), the glossopharyngeal nerve (CN IX), and the vagus nerve (CN X). (Copyright David Klemm, Faculty and Curriculum Support [FACS], Georgetown University Medical Center.)

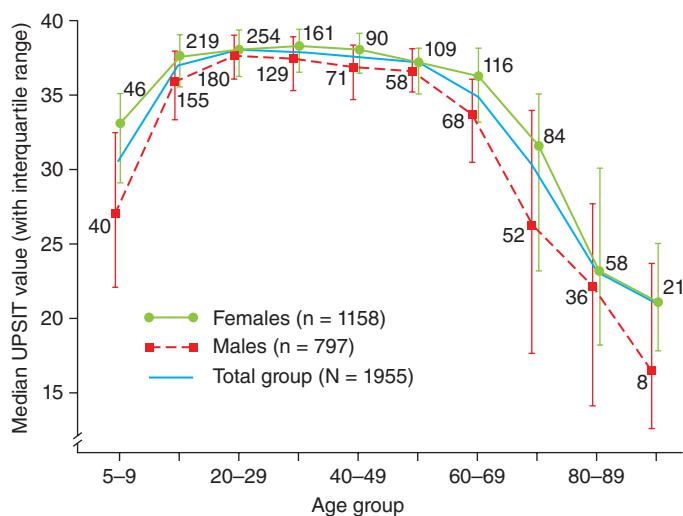


FIGURE 33-6 Scores on the University of Pennsylvania Smell Identification Test (UPSIT) as a function of subject age and sex. Numbers by each data point indicate sample sizes. Note that women identify odorants better than men at all ages. (RL Doty et al: Smell identification ability: Changes with age. *Science* 226:4681, 1984. Copyright © 1984 American Association for the Advancement of Science. Reprinted with permission from AAAS.)

TABLE 33-1 Disorders and Conditions Associated with Compromised Olfactory Function, as Measured by Olfactory Testing

Endocrine and Metabolic Conditions	Nasosinus Disorders	Viral, Bacterial, and Fungal Infections
Adrenal cortical insufficiency (Addison's disease)	Adenoid hypertrophy	Candidiasis
Chromatin-negative gonadal dysgenesis (Turner's syndrome)	Bacterial and viral upper respiratory infections	COVID-19
Cushing's syndrome	Laryngopharyngeal reflux disease	Hepatitis C
Diabetes	Rhinosinusitis/polyposis	Herpetic meningoencephalitis
Hypertension		Human immunodeficiency virus
Hypothyroidism		Legionnaires' disease
Idiopathic hypogonadotropic hypogonadism	Neurologic Diseases/Disorders	Leprosy (Hansen's disease)
Kallmann's syndrome	Alzheimer's disease	Lyme disease
Liver disease	Amyotrophic lateral sclerosis (ALS)	Poliomyelitis
Renal disease/kidney failure	Bell's palsy	Rhinosinusitis
Pregnancy	Degenerative ataxias	Upper respiratory infections
Pseudohypoparathyroidism	Down's syndrome	
Wilson's disease	Epilepsy	Other Disorders or Factors
Immune-Related Diseases	Facial paralysis	Alcoholism
Acute disseminated encephalomyelitis	Fibromyalgia	Bardet-Biedl syndrome
Allergic rhinitis	Frontotemporal lobe degeneration	Chemical exposure
Asthma	Guamanian ALS/Parkinson's disease/dementia syndrome	Congenital
Autoimmune pancreatitis	Head trauma	Iatrogenesis, including chemotherapy and radiation
Behçet's disease	Huntington's disease	Nutritional deficiencies
Churg-Strauss syndrome	Idiopathic inflammatory myopathies	Obesity
Cystic fibrosis	Korsakoff psychosis	Tobacco smoking
Fibromyalgia	Lubag disease	Toxic chemical exposures
Giant cell arteritis	Migraine	Vitamin B ₁₂ deficiency
Hereditary angioedema	Multi-infarct dementia	
Idiopathic inflammatory myopathies	Narcolepsy with cataplexy	
Inflammatory bowel diseases	Neoplasms, cranial/nasal	
Lupus	Orthostatic tremor	
Mikulicz's disease	Parkinson's disease	
Multiple sclerosis	Pick's disease	
Myasthenia gravis	Rapid eye movement behavioral sleep disorder	
Neuromyelitis optica	Stroke	
Pemphigus vulgaris		Psychiatric-Related Diseases/Disorders
Psoriasis vulgaris	Psychiatric-Related Diseases/Disorders	Anorexia nervosa
Rheumatoid arthritis		Asperger's syndrome
Sjögren's syndrome		Attention deficit/hyperactivity disorder
Systemic sclerosis (scleroderma)		Depression
Wegener's granulomatosis		Obsessive compulsive disorder
		Panic disorder
		Posttraumatic stress disorder
		Psychopathy
		Schizophrenia
		Seasonal affective disorder
		22q11 deletion syndrome

Note: These disease/disorder classifications are not necessarily mutually exclusive.

that food has little flavor, a problem that can result in nutritional disturbances. This also helps to explain why a disproportionate number of elderly die in accidental gas poisonings. A relatively complete listing of conditions and disorders that have been associated with olfactory dysfunction is presented in **Table 33-1**.

Aside from aging, the three most common identifiable causes of long-lasting or permanent smell loss seen in the clinic are, in order of frequency, severe upper respiratory infections, head trauma, and chronic rhinosinusitis. The physiologic basis for most head trauma-related losses is the shearing and subsequent scarring of the olfactory fila as they pass from the nasal cavity into the brain cavity. The cribriform plate does not have to be fractured or show pathology for smell loss to be present. Severity of trauma, as indexed by a poor Glasgow Coma Scale score on presentation and the length of posttraumatic amnesia, is associated with higher risk of olfactory impairment. Less than 10% of posttraumatic anosmic patients will recover age-related normal function over time. This increases to nearly 25% of those with less-than-total

loss. Respiratory infections, such as those associated with the common cold, influenza, pneumonia, HIV, and COVID-19 can directly and permanently damage the olfactory epithelium, decreasing receptor cell number, damaging cilia on remaining receptor cells, and inducing the replacement of sensory epithelium with respiratory epithelium. The smell loss associated with chronic rhinosinusitis is related to disease severity, with most loss occurring in cases where rhinosinusitis and polyposis are both present. Smell loss is among the first signs of the SARS-CoV-2 infection responsible for COVID-19, a loss that is seemingly independent of nasal inflammation. Although in rhinosinusitis cases systemic glucocorticoid therapy can usually induce short-term functional improvement, it does not, on average, return smell test scores to normal, implying that chronic permanent neural loss is present and/or that short-term administration of systemic glucocorticoids does not completely mitigate the inflammation. It is well established that microinflammation in an otherwise seemingly normal epithelium can influence smell function.

A number of neurodegenerative diseases are accompanied by olfactory impairment, including PD, AD, Huntington's disease, parkinsonism-dementia complex of Guam, dementia with Lewy bodies (DLB), multiple system atrophy, corticobasal degeneration, frontotemporal dementia, and Down's syndrome; smell loss can also occur in idiopathic rapid eye movement (REM) behavioral sleep disorder (iRBD), as well as in multiple sclerosis (MS) related to lesions within olfaction-related structures. Olfactory impairment in PD often predates the clinical diagnosis by a number of years. In staged cases, studies of the sequence of formation of abnormal α -synuclein aggregates and Lewy bodies suggest that the olfactory bulbs may be, along with the dorsomotor nucleus of the vagus, the first site of neural damage in PD. In postmortem studies of patients with very mild "presymptomatic" signs of AD, poorer smell function has been associated with higher levels of AD-related pathology. Smell loss is more marked in patients with early clinical manifestations of DLB than in those with mild AD. Interestingly, smell loss is minimal or nonexistent in progressive supranuclear palsy and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-induced parkinsonism. The relative contributions of disease-specific pathology or differential damage to forebrain neuromodulator/neurotransmitter systems in explaining different degrees of olfactory dysfunction among the various neurodegenerative diseases are presently unknown.

The smell loss seen in iRBD is of the same magnitude as that found in PD. This is of particular interest because patients with iRBD frequently develop PD and hyposmia. REM behavior disorder is not only seen in its idiopathic form, but can also be associated with narcolepsy (Chap. 31). A study of narcoleptic patients with and without REM behavior disorder demonstrated that narcolepsy, independent of REM behavior disorder, was associated with impairments in olfactory function. Loss of hypothalamic neurons expressing orexin (also known as hypocretin) neuropeptides is believed to be responsible for narcolepsy and cataplexy. Orexin-containing neurons project throughout the entire olfactory system (from the olfactory epithelium to the olfactory cortex), and damage to these projections may be one underlying mechanism for impaired olfactory performance in narcoleptic patients. Administration of intranasal orexin A (hypocretin-1) improved olfactory function, supporting the notion that mild olfactory impairment is not only a primary feature of narcolepsy with cataplexy, but that orexin deficiency may be directly responsible for the loss of smell in this condition.

■ DISORDERS OF TASTE

The majority of patients who present with taste dysfunction exhibit olfactory, not taste, loss. This is because most flavors attributed to taste actually depend on retronasal stimulation of the olfactory receptors during deglutition. As noted earlier, taste buds only mediate basic tastes such as sweet, sour, bitter, salty, and umami. Significant impairment of whole-mouth gustatory function is rare outside of generalized metabolic disturbances or systemic use of some medications, because taste bud regeneration occurs and peripheral damage alone would require the involvement of multiple CN pathways. Taste function can be influenced by age, diet, smoking behavior, use of medications, and other subject-related factors including (1) the release of foul-tasting materials from the oral cavity from oral medical conditions (e.g., gingivitis, purulent sialadenitis) or appliances; (2) transport problems of tastants to the taste buds (e.g., drying, infections, or inflammatory conditions of the orolingual mucosa), (3) damage to the taste buds themselves (e.g., local trauma, invasive carcinomas), (4) damage to the neural pathways innervating the taste buds (e.g., middle ear infections), (5) damage to central structures (e.g., multiple sclerosis, tumor, epilepsy, stroke), and (6) systemic disturbances of metabolism (e.g., diabetes, thyroid disease, medications).

Unlike CN VII, CN IX is relatively protected along its path, although iatrogenic interventions such as tonsillectomy, bronchoscopy, laryngoscopy, endotracheal intubation, and radiation therapy can result in selective injury. CN VII damage commonly results from mastoidectomy, tympanoplasty, and stapedectomy, in some cases inducing persistent metallic sensations. Bell's palsy (Chap. 441) is one of the most common causes of CN VII injury that results in taste disturbance. On

rare occasions, migraine (Chap. 430) is associated with a gustatory prodrome or aura, and in some cases, tastants can trigger a migraine attack. Interestingly, dysgeusia occurs in some cases of *burning mouth syndrome* (also termed *glossodynia* or *glossalgia*), as does dry mouth and thirst. Burning mouth syndrome is likely associated with dysfunction of the trigeminal nerve (CN V). Some of the etiologies suggested for this poorly understood syndrome are amenable to treatment, including (1) nutritional deficiencies (e.g., iron, folic acid, B vitamins, zinc), (2) diabetes mellitus (possibly predisposing to oral candidiasis), (3) denture allergy, (4) mechanical irritation from dentures or oral devices, (5) repetitive movements of the mouth (e.g., tongue thrusting, teeth grinding, jaw clenching), (6) tongue ischemia as a result of temporal arteritis, (7) periodontal disease, (8) reflux esophagitis, and (9) geographic tongue.

Although both taste and smell can be adversely influenced by drugs, taste alterations are more common. Indeed, >250 medications have been reported to alter the ability to taste. Major offenders include antineoplastic agents, antirheumatic drugs, antibiotics, and blood pressure medications. Terbinafine, a commonly used antifungal, has been linked to taste disturbance lasting up to 3 years. In a recent controlled trial, nearly two-thirds of individuals taking eszopiclone (Lunesta) for insomnia experienced a bitter dysgeusia that was stronger in women, systematically related to the time since drug administration, and positively correlated with both blood and saliva levels of the drug. Intranasal use of nasal gels and sprays containing zinc, which are common over-the-counter prophylactics for upper respiratory viral infections, has been implicated in loss of smell function. Whether their efficacy in preventing such infections, which are the most common cause of anosmia and hyposmia, outweighs their potential detriment to smell function requires study. Dysgeusia occurs commonly in the context of drugs used to treat or minimize symptoms of cancer, with a weighted prevalence from 56% to 76% depending on the type of cancer treatment. Attempts to prevent taste problems from such drugs using prophylactic zinc sulfate or amifostine have proven to be minimally beneficial. Although antiepileptic medications are occasionally used to treat smell or taste disturbances, the use of topiramate has been reported to result in a reversible loss of an ability to detect and recognize tastes and odors during treatment.

As with olfaction, a number of systemic disorders can affect taste. These include, but are not limited to, chronic renal failure, end-stage liver disease, vitamin and mineral deficiencies, diabetes mellitus, and hypothyroidism. In diabetes, there appears to be a progressive loss of taste beginning with glucose and then extending to other sweeteners, salty stimuli, and then all stimuli. Psychiatric conditions can be associated with chemosensory alterations (e.g., depression, schizophrenia, bulimia). A recent review of tactile, gustatory, and olfactory hallucinations demonstrated that no one type of hallucinatory experience is pathognomonic to any given diagnosis.

Pregnancy is a unique condition with regard to taste function. There appears to be an increase in dislike and intensity of bitter tastes during the first trimester that may help to ensure that pregnant women avoid poisons during a critical phase of fetal development. Similarly, a relative increase in the preference for salt and bitter in the second and third trimesters may support the ingestion of much needed electrolytes to expand fluid volume and support a varied diet.

■ CLINICAL EVALUATION

In most cases, a careful clinical history will establish the probable etiology of a chemosensory problem, including questions about its nature, onset, duration, and pattern of fluctuations. *Sudden loss* suggests the possibility of head trauma, ischemia, infection, or a psychiatric condition. *Gradual loss* can reflect the development of a progressive obstructive lesion, although gradual loss can also follow head trauma. *Intermittent loss* suggests the likelihood of an inflammatory process. The patient should be asked about potential precipitating events, such as cold or flu infections, prior to symptom onset, because these often go underappreciated. Information regarding head trauma, smoking habits, drug and alcohol abuse (e.g., intranasal cocaine, chronic alcoholism), exposures to pesticides and other toxic agents, and medical

interventions is also informative. A determination of all the medications that the patient was taking before and at the time of symptom onset is important, because many can cause chemosensory disturbances. Comorbid medical conditions associated with smell impairment, such as renal failure, liver disease, hypothyroidism, diabetes, or dementia, should be assessed. Delayed puberty in association with anosmia (with or without midline craniofacial abnormalities, deafness, and renal anomalies) suggests the possibility of Kallmann's syndrome. Recollection of epistaxis, discharge (clear, purulent, or bloody), nasal obstruction, allergies, and somatic symptoms, including headache or irritation, may have localizing value. Questions related to memory, parkinsonian symptoms, and seizure activity (e.g., automatisms, blackouts, auras, *déjà vu*) should be posed. Pending litigation and the possibility of malingering should be considered. Modern forced-choice olfactory tests can detect malingering from improbable responses.

Neurologic and otorhinolaryngologic (ORL) examinations, along with appropriate brain and nasosinus imaging, aid in the evaluation of patients with olfactory or gustatory complaints. The neural evaluation should focus on CN function, with particular attention to possible skull base and intracranial lesions. Visual acuity, field, and optic disc examinations aid in detection of intracranial mass lesions that produce raised intracranial pressure (papilledema) and optic atrophy. Foster Kennedy syndrome refers to raised intracranial pressure plus a compressive optic neuropathy; typical causes are olfactory groove meningiomas or other frontal lobe tumors. The ORL examination should thoroughly assess the intranasal architecture and mucosal surfaces. Polyps, masses, and adhesions of the turbinates to the septum may compromise the flow of air to the olfactory receptors, because less than a fifth of the inspired air traverses the olfactory cleft in the unobstructed state. Blood tests may be helpful to identify such conditions as diabetes, infection, heavy metal exposure, nutritional deficiency (e.g., vitamin B₆ or B₁₂), allergy, and thyroid, liver, and kidney disease.

As with other sensory disorders, quantitative sensory testing is advised. Self-reports of patients can be misleading, and a number of patients who complain of chemosensory dysfunction have normal function for their age and gender. Quantitative smell and taste testing provides objective information for worker's compensation and other legal claims, as well as a way to accurately assess the effects of treatment interventions. A number of standardized olfactory and taste tests are commercially available. The most widely used olfactory test, the 40-item University of Pennsylvania Smell Identification Test (UPSIT), uses norms based on nearly 4000 normal subjects. A determination is made of both absolute dysfunction (i.e., mild loss, moderate loss, severe loss, total loss, probable malingering) and relative dysfunction (percentile rank for age and gender). Although electrophysiologic testing is available at some smell and taste centers (e.g., odor event-related potentials), they require complex stimulus presentation and recording equipment and rarely provide additional diagnostic information. With the exception of electrogustometers, commercially available taste tests have only recently become available. Most use filter paper strips or similar materials impregnated with tastants, so no stimulus preparation is required.

TREATMENT AND MANAGEMENT

Given the various mechanisms by which olfactory and gustatory disturbance can occur, management of patients tends to be condition-specific. For example, patients with hypothyroidism, diabetes, or infections often benefit from specific treatments to correct the underlying disease process that is adversely influencing chemoreception. For most patients who present primarily with obstructive/transport loss affecting the nasal and paranasal regions (e.g., allergic rhinitis, polyposis, intranasal neoplasms, nasal deviations), medical and/or surgical intervention is often beneficial. Antifungal and antibiotic treatments may reverse taste problems secondary to candidiasis or other oral infections. Chlorhexidine mouthwash mitigates some salty or bitter dysgeusias, conceivably as a result of its strong positive charge. Excessive dryness of the oral mucosa is a problem with many medications and conditions, and artificial saliva (e.g., Xerolube) or oral pilocarpine treatments may prove beneficial. Other methods to improve salivary flow include the

use of mints, lozenges, or sugarless gum. Flavor enhancers may make food more palatable (e.g., monosodium glutamate), but caution is advised to avoid overusing ingredients containing sodium or sugar, particularly in circumstances when a patient also has underlying hypertension or diabetes. Medications that induce distortions of taste can often be discontinued and replaced with other types of medications or modes of therapy. As mentioned earlier, pharmacologic agents result in taste disturbances much more frequently than smell disturbances. It is important to note, however, that many drug-related effects are long lasting and not reversed by short-term drug discontinuance.

A study of endoscopic sinus surgery in patients with chronic rhinosinusitis and hyposmia revealed that patients with severe olfactory dysfunction prior to the surgery had a more dramatic and sustained improvement over time compared to patients with more mild olfactory dysfunction prior to intervention. In the case of intranasal and sinus-related inflammatory conditions, such as seen with allergy, viruses, and traumas, the use of intranasal or systemic glucocorticoids may also be helpful. One common approach is to use a tapering course of oral prednisone. Topical intranasal administration of glucocorticoids was found to be less effective in general than systemic administration; however, the effects of different nasal administration techniques were not analyzed. For example, intranasal glucocorticoids are more effective if administered in the Moffett's position (head in the inverted position such as over the edge of the bed with the bridge of the nose perpendicular to the floor). After head trauma, an initial trial of glucocorticoids may help to reduce local edema and the potential deleterious deposition of scar tissue around olfactory fila at the level of the cribriform plate.

Treatments are limited for patients with chemosensory loss or primary injury to neural pathways. Nonetheless, spontaneous recovery can occur. In a follow-up study of 542 patients presenting to our center with smell loss from a variety of causes, modest improvement occurred over an average time period of 4 years in about half of the participants. However, only 11% of the anosmic and 23% of the hyposmic patients regained normal age-related function. Interestingly, the amount of dysfunction at the time of presentation, not etiology, was the best predictor of prognosis. Other predictors were age and the duration of dysfunction prior to initial testing.

Several studies have reported that patients with hyposmia may benefit from repeated smelling of odors over the course of weeks or months, although it remains to be determined how much improvement, if any, occurs over that known to occur spontaneously. The usual paradigm is to smell odors such as eucalyptol, citronella, eugenol, and phenyl ethyl alcohol before going to bed and immediately upon awakening each day. The rationale for such an approach comes from animal studies demonstrating that prolonged exposure to odorants can induce increased neural activity within the olfactory bulb. There is also limited evidence that α-lipoic acid (400 mg/d), an essential cofactor for many enzyme complexes with possible antioxidant effects, may be beneficial in mitigating smell loss following viral infection of the upper respiratory tract. However, double-blind studies are needed to confirm this observation. α-Lipoic acid has also been suggested to be useful in some cases of hypogesia and burning mouth syndrome.

The use of zinc and vitamin A in treating olfactory disturbances is controversial, and there does not appear to be much benefit beyond replenishing established deficiencies. However, zinc has been shown to improve taste function secondary to hepatic deficiencies, and retinoids (bioactive vitamin A derivatives) are known to play an essential role in the survival of olfactory neurons. One protocol in which zinc was infused with chemotherapy treatments suggested a possible protective effect against developing taste impairment. Diseases of the alimentary tract can not only influence chemoreceptive function but also occasionally influence vitamin B₁₂ absorption. This can result in a relative deficiency of vitamin B₁₂, theoretically contributing to olfactory nerve disturbance. Vitamin B₂ (riboflavin) and magnesium supplements are reported in the alternative literature to aid in the management of migraine that, in turn, may be associated with smell dysfunction. Because vitamin D deficiency is a cofactor of chemotherapy-induced mucocutaneous toxicity and dysgeusia, adding vitamin D₃, 1000–2000

units per day, may benefit some patients with smell and taste complaints during or following chemotherapy.

A number of medications have reportedly been used with success in ameliorating olfactory symptoms, although strong scientific evidence for efficacy is generally lacking. A report that theophylline improved smell function was uncontrolled and failed to account for the fact that some meaningful improvement occurs without treatment; indeed, the percentage of responders was about the same (~50%) as that noted by others to show spontaneous improvement over a similar time period. Antiepileptics and some antidepressants (e.g., amitriptyline) have been used to treat dysosmias and smell distortions, particularly following head trauma. Ironically, amitriptyline is also frequently on the list of medications that can ultimately distort smell and taste function, possibly from its anticholinergic effects. One study suggested that the centrally acting acetylcholinesterase inhibitor donepezil in AD resulted in improvements on smell identification measures that correlated with overall clinician-based impressions of change in dementia severity scores.

Alternative therapies, such as acupuncture, meditation, cognitive-behavioral therapy, and yoga, can help patients manage uncomfortable experiences associated with chemosensory disturbance and oral pain syndromes and to cope with the psychosocial stressors surrounding the impairment. Additionally, modification of diet and eating habits is also important. By accentuating the other sensory experiences of a meal, such as food texture, aroma, temperature, and color, one can optimize the overall eating experience for a patient. In some cases, a flavor enhancer like monosodium glutamate (MSG) can be added to foods to increase palatability and encourage intake.

Proper oral and nasal hygiene and routine dental care are extremely important ways for patients to protect themselves from disorders of the mouth and nose that can ultimately result in chemosensory disturbance. Patients should be warned not to overcompensate for their taste loss by adding excessive amounts of sugar or salt. Smoking cessation and the discontinuance of oral tobacco use are essential in the management of any patient with smell and/or taste disturbance and should be repeatedly emphasized.

A major and often overlooked element of therapy comes from chemosensory testing itself. Confirmation or lack of conformation of loss is beneficial to patients who come to believe, in light of unsupportive family members and medical providers, that they may be “crazy.” In cases where the loss is minor, patients can be informed of the likelihood of a more positive prognosis. Importantly, quantitative testing places the patient’s problem into overall perspective. Thus, it is often therapeutic for an older person to know that, while his or her smell function is not what it used to be, it still falls above the average of his or her peer group. Without testing, many such patients are simply told that they are getting old and nothing can be done for them, leading in some cases to depression and decreased self-esteem.

FURTHER READING

- DEVANAND DP et al: Olfactory identification deficits are associated with increased mortality in a multiethnic urban community. *Ann Neurol* 78:401, 2015.
- DOTY RL: Olfaction in Parkinson’s disease and related disorders. *Neurobiol Dis* 46:527, 2012.
- DOTY RL et al: Taste function in early stage treated and untreated Parkinson’s disease. *J Neurol* 262:547, 2015.
- DOTY RL et al: Systemic diseases and disorders. *Handbook Clin Neurol* 164:361, 2019.
- DOTY RL et al: Treatments for smell and taste disorders: A critical review. *Handbook Clin Neurol* 164:455, 2019.
- FORNAZIERI MA et al: Adherence and efficacy of olfactory training as a treatment for persistent olfactory loss. *Am J Rhinol Allergy* 34:238, 2020.
- HAWKES CH, DOTY RL: *Smell and Taste Disorders*. Cambridge, Cambridge University Press, 2018.
- LIU G et al: Prevalence and risk factors of taste and smell impairment in a nationwide sample of the US population: A cross-sectional study. *BMJ Open* 6:e013246, 2016.

LONDON B et al: Predictors of prognosis in patients with olfactory disturbance. *Ann Neurol* 63:159, 2008.

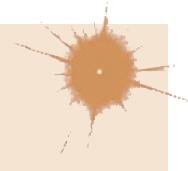
MOEIN ST et al: Smell dysfunction: A biomarker for COVID-19. *Int Forum Allergy Rhinol* 10:944, 2020.

PERRICONE C et al: Smell and autoimmunity: A comprehensive review. *Clin Rev Allergy Immunol* 45:87, 2013.

34

Disorders of Hearing

Anil K. Lalwani



Hearing loss can present at any age and is one of the most common sensory disorders in humans. Nearly 10% of the adult population has some hearing loss, and one-third of individuals age >65 years have a hearing loss of sufficient magnitude to require a hearing aid.

PHYSIOLOGY OF HEARING

The function of the external and middle ear is to amplify sound to facilitate conversion of the mechanical energy of the sound wave into an electrical signal by the inner-ear hair cells, a process called mechanotransduction (Fig. 34-1). Sound waves enter the external auditory canal and set the tympanic membrane (eardrum) in motion, which in turn moves the malleus, incus, and stapes of the middle ear. Movement of the footplate of the stapes causes pressure changes in the fluid-filled inner ear, eliciting a traveling wave in the basilar membrane of the cochlea. The tympanic membrane and the ossicular chain in the middle ear serve as an impedance-matching mechanism, improving the efficiency of energy transfer from air to the fluid-filled inner ear. In its absence, nearly 99.9% of the acoustical energy would be reflected and thus not heard. Instead, the eardrum and the ossicles boost the sound energy nearly 200-fold by the time it reaches the inner ear.

Within the cochlea of the inner ear, there are two types of hair cells that aid in hearing: inner and outer. The inner and outer hair cells of the organ of Corti have different innervation patterns, but both are mechanoreceptors; they detect the mechanical energy of the acoustic signal and aid its conversion to an electrical signal that travels by the auditory nerve. The afferent innervation relates principally to the inner hair cells while the efferent innervation relates principally to the outer hair cells. The outer hair cells outnumber the inner hair cells by nearly 6:1 (20,000 vs 3500). The motility of the outer hair cells alters the micromechanics of the inner hair cells, creating a cochlear amplifier, which explains the exquisite sensitivity and frequency selectivity of the cochlea.

Stereocilia of the hair cells of the organ of Corti, which rests on the basilar membrane, are in contact with the tectorial membrane and are deformed by the traveling wave. The deformation stretches tiny filamentous connections (tip links) between stereocilia, leading to opening of ion channels, influx of potassium, and hair cell depolarization and subsequent neurotransmission. A point of maximal displacement of the basilar membrane is determined by the frequency of the stimulating tone. High-frequency tones cause maximal displacement of the basilar membrane near the base of the cochlea, whereas for low-frequency sounds, the point of maximal displacement is toward the apex of the cochlea.

Beginning in the cochlea, the frequency specificity is maintained at each point of the central auditory pathway: dorsal and ventral cochlear nuclei, trapezoid body, superior olive complex, lateral lemniscus, inferior colliculus, medial geniculate body, and auditory cortex. At low frequencies, individual auditory nerve fibers can respond more or less synchronously with the stimulating tone. At higher frequencies, phase-locking occurs so that neurons alternate in response to particular phases of the cycle of the sound wave. Intensity is encoded by the

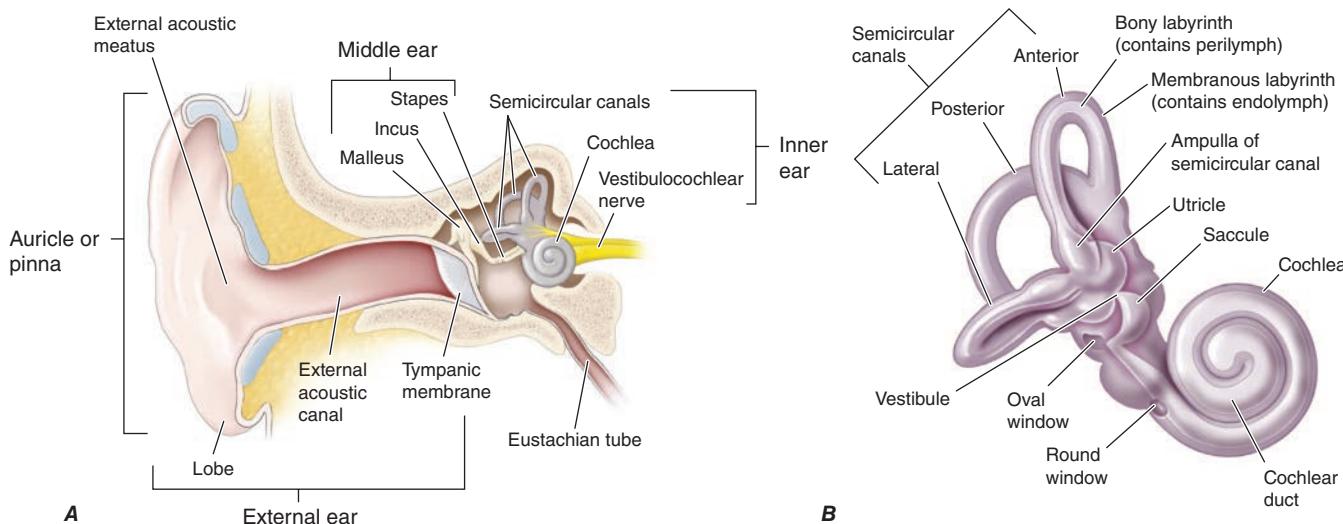


FIGURE 34-1 Ear anatomy. **A.** Drawing of modified coronal section through external ear and temporal bone, with structures of the middle and inner ear demonstrated. **B.** High-resolution view of inner ear.

amount of neural activity in individual neurons, the number of neurons that are active, and the specific neurons that are activated.

There is evidence that the right and left ears as well as the central nervous system may process speech asymmetrically. Generally, a sound is processed symmetrically from the peripheral to the central auditory system. However, a “right ear advantage” exists for dichotic listening tasks, in which subjects are asked to report on competing sounds presented to each ear. In most individuals, a perceptual right ear advantage for consonant-vowel syllables, stop consonants, and words also exists. Similarly, whereas central auditory processing for sounds is symmetric with minimal lateral specialization for the most part, speech processing is lateralized. There is specialization of the left auditory cortex for speech recognition and production, and of the right hemisphere for emotional and tonal aspects of speech. Left hemisphere dominance for speech is found in 95–98% of right-handed persons and 70–80% of left-handed persons.

DISORDERS OF THE SENSE OF HEARING

Hearing loss can result from disorders of the auricle, external auditory canal, middle ear, inner ear, or central auditory pathways (Fig. 34-2). In general, lesions in the auricle, external auditory canal, or middle ear that impede the transmission of sound from the external environment to the inner ear cause conductive hearing loss, whereas lesions that impair mechanotransduction in the inner ear or transmission of the electrical signal along the eighth nerve to the brain cause sensorineural hearing loss.

Conductive Hearing Loss The external ear, the external auditory canal, and the middle-ear apparatus are designed to collect and amplify sound and efficiently transfer the mechanical energy of the sound wave to the fluid-filled cochlea. Factors that obstruct the transmission of sound or dampen the acoustic energy result in conductive hearing loss. Conductive hearing loss can occur from obstruction of the external auditory canal by cerumen, debris, and foreign bodies; swelling of the lining of the canal; atresia or neoplasms of the canal; perforations of the tympanic membrane; disruption of the ossicular chain, as occurs with necrosis of the long process of the incus in trauma or infection; otosclerosis; or fluid, scarring, or neoplasms in the middle ear. Rarely, inner-ear malformations or pathologies that create a “third window” in the inner ear such as superior semicircular canal dehiscence, lateral semicircular canal dysplasia, incomplete partition of the inner ear, and large vestibular aqueduct, are also associated with conductive hearing loss. This pathologic third window is associated with loss of mechanical energy associated with the sound wave leading to conductive hearing loss (see below).

Eustachian tube dysfunction is extremely common in adults and may predispose to acute otitis media (AOM) or serous otitis media

(SOM). Recently, Eustachian tube balloon dilation has been shown to relieve acquired inflammatory obstruction of the Eustachian tube orifice and improve symptoms due to Eustachian tube dysfunction. Trauma, AOM, and chronic otitis media are the usual factors responsible for tympanic membrane perforation. While small perforations often heal spontaneously, larger defects usually require surgical intervention. Tympanoplasty is highly effective (>90%) in the repair of tympanic membrane perforations. Otoscopy is usually sufficient to diagnose AOM, SOM, chronic otitis media, cerumen impaction, tympanic membrane perforation, and Eustachian tube dysfunction; tympanometry and Eustachian tube function testing can be useful to confirm the clinical suspicion of these conditions.

Cholesteatoma, a benign tumor composed of stratified squamous epithelium in the middle ear or mastoid, occurs frequently in adults, often in the setting of severe Eustachian tube dysfunction. This is a slowly growing lesion that destroys bone and normal ear tissue. Theories of pathogenesis include traumatic immigration and invasion of squamous epithelium through a retraction pocket of the tympanic membrane, implantation of squamous epithelia in the middle ear through a perforation or surgery, and metaplasia following chronic infection and irritation. A chronically draining ear that fails to respond to appropriate antibiotic therapy should raise suspicion of a cholesteatoma. On examination, there is often a perforation of the tympanic membrane filled with cheesy white squamous debris. The presence of an aural polyp obscuring the tympanic membrane is highly suggestive of an underlying cholesteatoma. Conductive hearing loss secondary to ossicular erosion is common. Bony destruction visualized on CT of the temporal bone is also highly suggestive of cholesteatoma. Surgery is required to remove this destructive process and reconstruct the ossicles.

Conductive hearing loss with a normal ear canal and intact tympanic membrane suggests either ossicular pathology or the presence of a “third window” in the inner ear (see below). Fixation of the stapes from *otosclerosis* is a common cause of low-frequency conductive hearing loss. It occurs equally in men and women and is inherited as an autosomal dominant trait with incomplete penetrance; in some cases, it may be a manifestation of osteogenesis imperfecta. Hearing impairment usually presents between the late teens and the forties. In women, the otosclerotic process is accelerated during pregnancy, and the hearing loss is often first noticeable at this time. A hearing aid or a simple outpatient surgical procedure (stapedectomy) can provide excellent auditory rehabilitation. Extension of otosclerosis beyond the stapes footplate to involve the cochlea (cochlear otosclerosis) can lead to mixed or sensorineural hearing loss. Fluoride therapy to prevent hearing loss from cochlear otosclerosis is of uncertain value.

Disorders that lead to the formation of a pathologic “third window” in the inner ear can be associated with conductive hearing loss. There

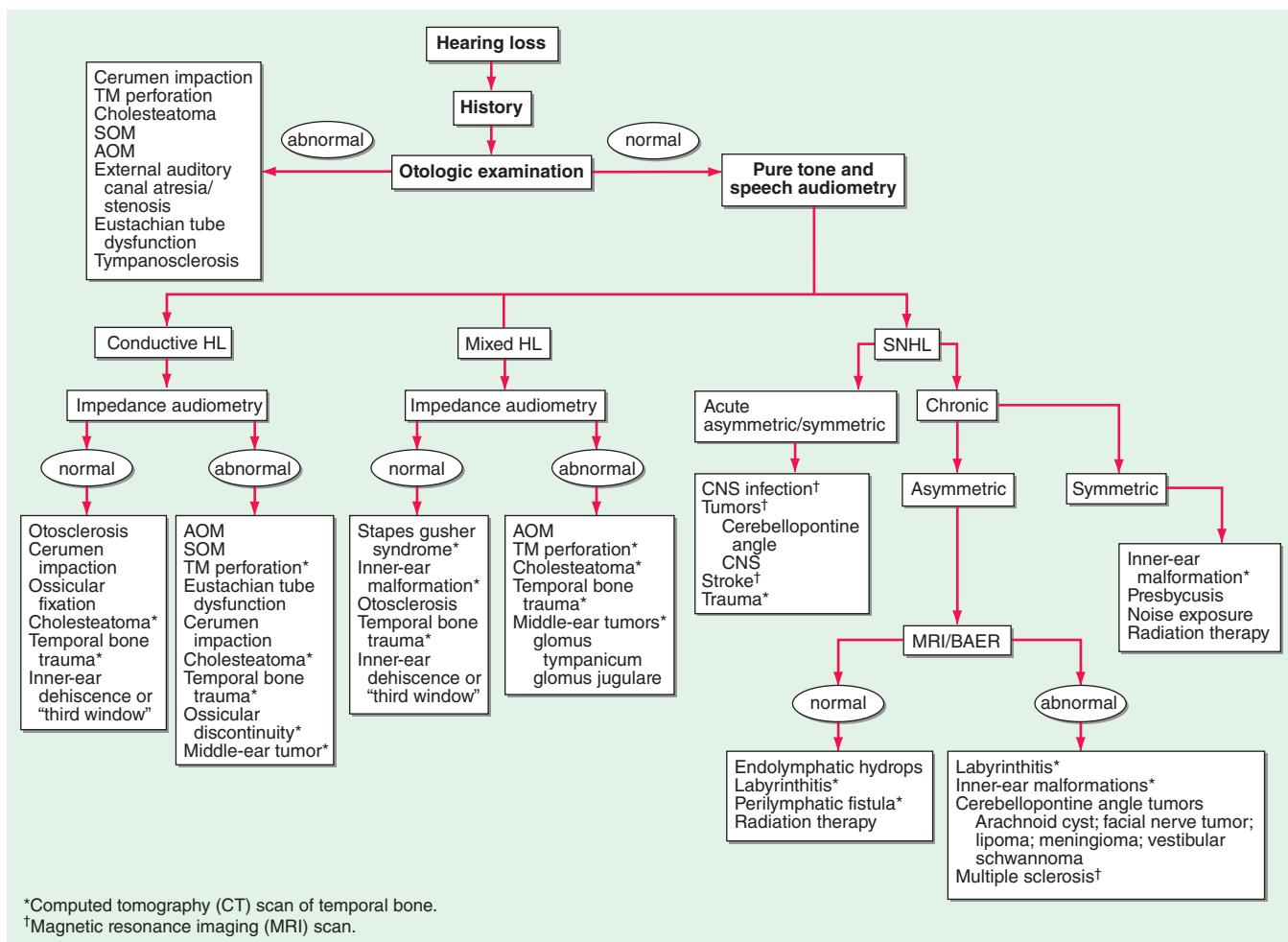


FIGURE 34-2 An algorithm for the approach to hearing loss. AOM, acute otitis media; BAER, brainstem auditory-evoked response; CNS, central nervous system; HL, hearing loss; SNHL, sensorineural hearing loss; SOM, serous otitis media; TM, tympanic membrane.

are normally two major openings, or windows, that connect the inner ear with the middle ear and serve as conduits for transmission of sound; these are, respectively, the oval and round windows. A third window is formed where the normally hard otic bone surrounding the inner ear is eroded; dissipation of the acoustic energy at the third window is responsible for the “inner-ear conductive hearing loss.” The superior semicircular canal dehiscence syndrome resulting from erosion of the otic bone over the superior circular canal can present with conductive hearing loss that mimics otosclerosis. A common symptom is vertigo evoked by loud sounds (Tullio phenomenon), by Valsalva maneuvers that change middle-ear pressure, or by applying positive pressure on the tragus (the cartilage anterior to the external opening of the ear canal). Patients with this syndrome also complain of fullness of the ear, pulsatile tinnitus, and being able to hear the movement of their eyes and neck. A large jugular bulb or jugular bulb diverticulum can create a “third window” by eroding into the vestibular aqueduct or posterior semicircular canal; the symptoms are similar to those of the superior semicircular canal dehiscence syndrome. Other inner-ear malformations such as lateral semicircular canal dysplasia, large vestibular aqueduct, or incomplete partition seen in stapes gusher syndrome can also be associated with inner-ear conductive hearing loss as a result of the third window. Low activation threshold on the vestibular-evoked myogenic potential test (VEMP test, see below) and inner-ear erosion on CT are diagnostic. Recalcitrant vertigo and dizziness may respond to surgical repair of the dehiscence.

Sensorineural Hearing Loss Sensorineural hearing loss results from either damage to the mechanotransduction apparatus of the cochlea or disruption of the electrical conduction pathway from the inner ear to the brain. Thus, injury to hair cells, supporting cells,

auditory neurons, or the central auditory pathway can cause sensorineural hearing loss. Damage to the hair cells of the organ of Corti may be caused by intense noise, viral infections, ototoxic drugs (e.g., salicylates, quinine and its synthetic analogues, aminoglycoside antibiotics, loop diuretics such as furosemide and ethacrynic acid, and cancer chemotherapeutic agents such as cisplatin), fractures of the temporal bone, meningitis, cochlear otosclerosis (see above), Ménière’s disease, and aging. Congenital malformations of the inner ear may be the cause of hearing loss in some adults. Genetic predisposition alone or in concert with environmental exposures may also be responsible (see below).

Noise-Induced Hearing Loss Exposure to loud noise, either a short burst or over a more prolonged period of time, can lead to noise-induced hearing loss. Acute exposure to noise can lead to either temporary or permanent threshold shifts, depending on the intensity and duration of sound, due to hair cell injury and/or death. Typically, with permanent hearing loss there is a “noise notch” with elevated hearing thresholds at 3000–4000 Hz. More recently, loud noise exposure has also been associated with “hidden hearing loss”—hidden, because routine audiometry shows the pure tone hearing to be normal. Patients usually complain of not being able to hear clearly and are more bothered by the presence of background noise. In contrast to hair cell loss, hidden hearing loss is thought to be due to loss of auditory synapses on hair cells following noise exposure. In an increasingly noisy world, avoiding acoustic trauma with earplugs or earmuffs is highly recommended to prevent noise-induced or hidden hearing loss.

Presbycusis (age-associated hearing loss) is the most common cause of sensorineural hearing loss in adults. It is estimated to affect over half of adults aged >75 years in the United States, a population that is expected to double in size over the next 40 years. In the early stages, it is

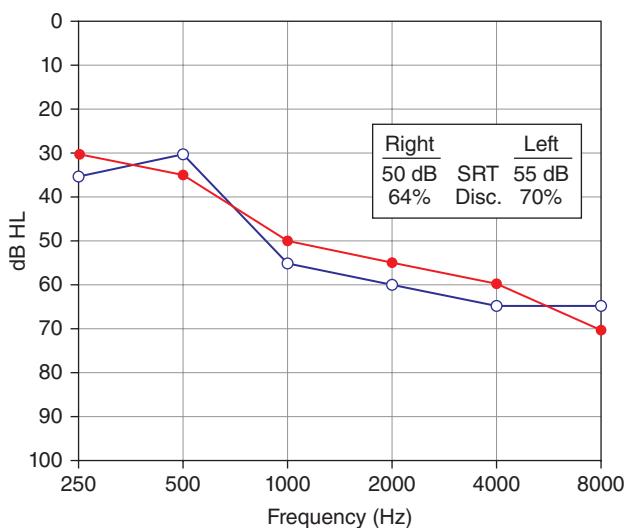


FIGURE 34-3 Presbycusis or age-related hearing loss. The audiogram shows a moderate to severe downsloping sensorineural hearing loss typical of presbycusis. The loss of high-frequency hearing is associated with a decreased speech discrimination score; consequently, patients complain of lack of clarity of hearing, especially in a noisy background. HL, hearing threshold level; SRT, speech reception threshold.

characterized by symmetric, gentle to sharply sloping, high-frequency hearing loss (Fig. 34-3). With progression, the hearing loss involves all frequencies. More importantly, the hearing impairment is associated with significant loss in clarity. There is a loss of discrimination for phonemes, recruitment (abnormal growth of loudness), and particular difficulty in understanding speech in noisy environments such as at restaurants and social events. Poor hearing is also associated with an increased incidence of cognitive impairment, rate of cognitive decline, and falls. In the elderly, left untreated, hearing loss leads to diminished quality of life, and has been shown to increase overall morbidity and mortality through falls and accidents. Hearing aids are helpful in enhancing the signal-to-noise ratio by amplifying sounds that are close to the listener. Hearing aid use has been shown to reduce cognitive decline and risk of falls. Although hearing aids are able to amplify sounds, they cannot restore the clarity of hearing. Thus, amplification with hearing aids may provide only limited rehabilitation once the word recognition score deteriorates below 50%. Cochlear implants are the treatment of choice when hearing aids prove inadequate, even when hearing loss is incomplete (see below).

Ménière's disease is characterized by episodic vertigo, fluctuating sensorineural hearing loss, tinnitus, and aural fullness. An absence of vertigo is inconsistent with the diagnosis of *Ménière's disease*, and the presence of fluctuating sensorineural hearing loss, tinnitus, and fullness without vertigo is more suggestive of cochlear hydrops. Tinnitus and/or deafness may be absent during the initial attacks of vertigo, but invariably appear as the disease progresses and increases in severity during acute attacks. The annual incidence of *Ménière's disease* is 0.5–7.5 per 1000; onset is most frequently in the fifth decade of life but may also occur in young adults or the elderly. Histologically, there is distention of the endolymphatic system (endolymphatic hydrops) leading to degeneration of vestibular and cochlear hair cells. This may result from endolymphatic sac dysfunction secondary to infection, trauma, autoimmune disease, inflammatory causes, or tumor; an idiopathic etiology constitutes the largest category and is most accurately referred to as *Ménière's disease*. Endolymphatic sac tumors, often associated with von Hippel Lindau disease, may clinically mimic *Ménière's disease*. Although any pattern of hearing loss can be observed, typically, low-frequency, unilateral sensorineural hearing impairment is present. An abnormal VEMP test (see below) may be helpful in detecting *Ménière's disease* in a clinically unaffected contralateral ear. MRI should be obtained to exclude retrocochlear pathology such as a cerebellopontine angle tumor, endolymphatic sac tumor, or demyelinating disorder. Therapy is directed toward the control of vertigo. A 2-g/d low-salt diet

is the mainstay of treatment for control of rotatory vertigo. Diuretics, a short course of oral glucocorticoids, intratympanic glucocorticoids, or intratympanic gentamicin may also be useful adjuncts in recalcitrant cases. Surgical therapy of vertigo is reserved for unresponsive cases and includes endolymphatic sac decompression, labyrinthectomy, and vestibular nerve section. Both labyrinthectomy and vestibular nerve section abolish rotatory vertigo in >90% of cases. Unfortunately, there is no effective therapy for hearing loss, tinnitus, or aural fullness from *Ménière's disease*.

Sensorineural hearing loss may also result from any neoplastic, vascular, demyelinating, infectious, degenerative disease, or trauma affecting the central auditory pathways. Characteristically, in hearing loss due to central nervous system pathology, a reduction in clarity of hearing and speech comprehension is much greater than the loss of the ability to hear pure tone. Auditory testing is consistent with an auditory neuropathy; normal otoacoustic emissions (OAEs) and an abnormal auditory brainstem response (ABR) are typical (see below). Hearing loss can accompany hereditary sensorimotor neuropathies and inherited disorders of myelin. Tumors of the cerebellopontine angle such as vestibular schwannoma and meningioma (Chap. 90) usually present with asymmetric sensorineural hearing loss with greater deterioration of speech understanding than pure tone hearing. Multiple sclerosis (Chap. 444) may present with acute unilateral or bilateral hearing loss; typically, pure tone testing remains relatively stable while speech understanding fluctuates. Isolated labyrinthine infarction can present with acute hearing loss and vertigo due to a cerebrovascular accident involving the posterior circulation, usually the anterior inferior cerebellar artery; it may also be the heralding sign of impending catastrophic basilar artery infarction (Chap. 426). HIV (Chap. 202), which can produce both peripheral and central auditory system pathology, is another consideration in the evaluation of sensorineural hearing impairment.

A finding of conductive and sensorineural hearing loss in combination is termed *mixed hearing loss*. Mixed hearing losses can result from pathology of both the middle and inner ear, as can occur in otosclerosis involving the ossicles and the cochlea, head trauma, chronic otitis media, cholesteatoma, middle-ear tumors, and some inner-ear malformations.

Trauma resulting in temporal bone fractures may be associated with conductive, sensorineural, or mixed hearing loss. If the fracture spares the inner ear, there may simply be conductive hearing loss due to rupture of the tympanic membrane or disruption of the ossicular chain. These abnormalities can be surgically corrected. Profound hearing loss and severe vertigo are associated with temporal bone fractures involving the inner ear. A perilymphatic fistula associated with leakage of inner-ear fluid into the middle ear can occur and may require surgical repair. An associated facial nerve injury is not uncommon. CT is best suited to assess fracture of the traumatized temporal bone, evaluate the ear canal, and determine the integrity of the ossicular chain and involvement of the inner ear. Cerebrospinal fluid leaks that accompany temporal bone fractures are usually self-limited; the value of prophylactic antibiotics is uncertain.

Tinnitus Tinnitus is defined as the perception of a sound when there is no sound in the environment. It can have a buzzing, roaring, or ringing quality and may be pulsatile (synchronous with the heartbeat). Tinnitus is often associated with either a conductive or sensorineural hearing loss. The pathophysiology of tinnitus is not well understood. The cause of the tinnitus can usually be determined by finding the cause of the associated hearing loss. Tinnitus may be the first symptom of a serious condition such as a vestibular schwannoma. Pulsatile tinnitus requires evaluation of the vascular system of the head to exclude vascular tumors such as glomus jugulare tumors, aneurysms, dural arteriovenous fistulas, and stenotic arterial lesions; it may also occur with SOM, superior semicircular dehiscence, and inner-ear dehiscence. It is most commonly associated with some abnormality of the jugular bulb such as a large jugular bulb or jugular bulb diverticulum. In absence of demonstrated pathology on MRA/MRV or CT angiography, pulsatile tinnitus is usually attributed to turbulent venous blood flow through the transverse sinus, sigmoid sinus, and the jugular bulb.

■ GENETIC CAUSES OF HEARING LOSS

 More than half of childhood hearing impairment is thought to be hereditary; hereditary hearing impairment (HHI) can also manifest later in life. HHI may be classified as either nonsyndromic, when hearing loss is the only clinical abnormality, or syndromic, when hearing loss is associated with anomalies in other organ systems. Nearly two-thirds of HHIs are nonsyndromic. Between 70% and 80% of nonsyndromic HHI is inherited in an autosomal recessive manner and designated DFNB; another 15–20% is autosomal dominant (DFNA). Less than 5% is X-linked (DFNX) or maternally inherited via the mitochondria.

More than 150 loci harboring genes for nonsyndromic HHI have been mapped, with recessive loci outnumbering dominant ones; numerous genes have now been identified (**Table 34-1**). The hearing genes fall into the categories of structural proteins (*MYH9*, *MYO7A*, *MYO15*, *TECTA*, *DIAPH1*), transcription factors (*POU3F4*, *POU4F3*), ion channels (*KCNQ4*, *SLC26A4*), and gap junction proteins (*GJB2*, *GJB3*, *GJB6*). Several of these genes, including *GJB2*, *TECTA*, and *TMC1*, cause both autosomal dominant and recessive forms of nonsyndromic HHI. In general, the hearing loss associated with dominant genes has its onset in adolescence or adulthood, varies in severity, and progresses with age, whereas the hearing loss associated with recessive inheritance is congenital and profound. Connexin 26, a product of the *GJB2* gene, is particularly important because it is responsible for nearly 20% of all cases of childhood deafness; half of genetic deafness in children is *GJB2* related. Two frameshift mutations, 35delG and 167delT, account for >50% of the cases; however, screening for these two mutations alone is insufficient, and sequencing of the entire gene is required to fully capture *GJB2*-related recessive deafness. The 167delT mutation is highly prevalent in Ashkenazi Jews; ~1 in 1765 individuals in this population is homozygous and affected. *GJB2* hearing loss can also vary among the members of the same family, suggesting that other genes or factors influence the auditory phenotype. A single mutation in *GJB2* in combination with a single mutation in *GJB6* (connexin 30) can also lead to hearing loss and is an example of digenic inheritance of hearing loss.

In addition to *GJB2*, several other nonsyndromic genes are associated with hearing loss that progresses with age. The contribution of genetics to presbycusis is also becoming better understood and likely reflects a combination of genetic susceptibility impacted by environmental exposure to sound. Sensitivity to aminoglycoside ototoxicity can be maternally transmitted through a mitochondrial mutation. Susceptibility to noise-induced hearing loss may also be genetically determined.

There are >400 syndromic forms of hearing loss. These include Usher's syndrome (retinitis pigmentosa and hearing loss), Waardenburg's syndrome (pigmentary abnormality and hearing loss), Pendred's syndrome (thyroid organification defect and hearing loss), Alport's syndrome (renal disease and hearing loss), Jervell and Lange-Nielsen syndrome (prolonged QT interval and hearing loss), neurofibromatosis type 2 (bilateral acoustic schwannoma), and mitochondrial disorders (mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes [MELAS]; myoclonic epilepsy and ragged red fibers [MERRF]; and progressive external ophthalmoplegia [PEO]) (**Table 34-2**).

APPROACH TO THE PATIENT

Disorders of the Sense of Hearing

The goal in the evaluation of a patient with auditory complaints is to determine (1) the nature of the hearing impairment (conductive vs sensorineural vs mixed), (2) the severity of the impairment (mild, moderate, severe, or profound), (3) the anatomy of the impairment (external ear, middle ear, inner ear, or central auditory pathway), and (4) the etiology. The presence of signs and symptoms associated with hearing loss should be ascertained (**Table 34-3**). The history should elicit characteristics of the hearing loss, including the duration of deafness, unilateral versus bilateral involvement, nature of onset (sudden vs insidious), and rate of progression (rapid vs slow). Symptoms of tinnitus, vertigo, imbalance, aural fullness,

otorrhea, headache, facial nerve dysfunction, and head and neck paresthesias should be noted. Information regarding head trauma, exposure to ototoxins, occupational or recreational noise exposure, and family history of hearing impairment may also be important. A sudden onset of unilateral hearing loss, with or without tinnitus, may represent a viral infection of the inner ear, vestibular schwannoma, or a stroke. Patients with unilateral hearing loss (sensory or conductive) usually complain of reduced hearing, poor sound localization, and difficulty hearing clearly in the presence of background noise. Gradual progression of a hearing deficit is common with otosclerosis, noise-induced hearing loss, vestibular schwannoma, or Ménière's disease. Small vestibular schwannomas typically present with asymmetric hearing impairment, tinnitus, and imbalance (rarely vertigo); cranial neuropathy, in particular of the trigeminal or facial nerve, may accompany larger tumors. In addition to hearing loss, Ménière's disease may be associated with episodic vertigo, tinnitus, and aural fullness. Sound-induced vertigo, autophony, and being able to hear one's own neck or eye movement are highly suggestive of superior semicircular canal dehiscence. Hearing loss with otorrhea is most likely due to chronic otitis media or cholesteatoma.

Examination should include the auricle, external ear canal, and tympanic membrane. In the elderly, the external ear canal is often dry and fragile; it is preferable to clean cerumen with wall-mounted suction or cerumen loops and to avoid irrigation. Irrigation should also be avoided when a tympanic membrane perforation is present or the integrity of the eardrum cannot be established. In examining the eardrum, the topography of the tympanic membrane is more important than the presence or absence of the light reflex. In addition to the pars tensa (the lower two-thirds of the tympanic membrane), the pars flaccida (upper one-third of the tympanic membrane) above the short process of the malleus should also be examined for retraction pockets that may be evidence of chronic Eustachian tube dysfunction or cholesteatoma. Insufflation of the ear canal is necessary to assess tympanic membrane mobility and compliance. Careful inspection of the nose, nasopharynx, and upper respiratory tract is important. Unilateral serous effusion or unexplained otalgia should prompt a fiberoptic examination of the nasopharynx and larynx to exclude neoplasms. Cranial nerves should be evaluated with special attention to facial and trigeminal nerves, which are commonly affected with tumors involving the cerebellopontine angle.

The Rinne and Weber tuning fork tests, with a 512-Hz tuning fork, are used to screen for hearing loss, differentiate conductive from sensorineural hearing losses, and confirm the findings of audiologic evaluation. The Rinne test compares the ability to hear by air conduction with the ability to hear by bone conduction. The tines of a vibrating tuning fork are held near the opening of the external auditory canal, and then the stem is placed on the mastoid process; for direct contact, it may be placed on teeth or dentures. The patient is asked to indicate whether the tone is louder by air conduction or bone conduction. Normally, and in the presence of sensorineural hearing loss, a tone is heard louder by air conduction than by bone conduction; however, with conductive hearing loss of ≥30 dB (see "Audiologic Assessment," below), the bone-conduction stimulus is perceived as louder than the air-conduction stimulus. For the Weber test, the stem of a vibrating tuning fork is placed on the head in the midline and the patient is asked whether the tone is heard in both ears or better in one ear than in the other. With a unilateral conductive hearing loss, the tone is perceived in the affected ear. With a unilateral sensorineural hearing loss, the tone is perceived in the unaffected ear. A 5-dB difference in hearing between the two ears is required for lateralization.

■ LABORATORY ASSESSMENT OF HEARING

Audiologic Assessment The minimum audiologic assessment for hearing loss should include the measurement of pure tone air-conduction and bone-conduction thresholds, speech reception threshold, word recognition score, tympanometry, acoustic reflexes, and

TABLE 34-1 Hereditary Hearing Impairment Genes

DESIGNATION	GENE	FUNCTION	DESIGNATION	GENE	FUNCTION
Autosomal Dominant					Autosomal Recessive
DFNA1	<i>DIAPH1</i>	Cytoskeletal protein	DFNB1A	<i>GJB2</i>	Gap junction
DFNA2A	<i>KCNQ4</i>	Potassium channel	DFNB1B	<i>GJB6</i>	Gap junction
DFNA2B	<i>GJB3</i>	Gap junction	DFNB2	<i>MYO7A</i>	Cytoskeletal protein
DFNA2C	<i>IFNLR1</i>	Class II cytokine receptor	DFNB3	<i>MYO15A</i>	Cytoskeletal protein
DFNA3A	<i>GJB2</i>	Gap junction	DFNB4	<i>SLC26A4</i>	Chloride/iodide transporter
DFNA3B	<i>GJB6</i>	Gap junction	DFNB6	<i>TMIE</i>	Transmembrane protein
DFNA4A	<i>MYH14</i>	Class II nonmuscle myosin	DFNB7/B11	<i>TMC1</i>	Transmembrane protein
DFNA4B	<i>CEACAM16</i>	Cell adhesion molecule	DFNB8/10	<i>TPRSS3</i>	Transmembrane serine protease
DFNA5	<i>GSDME/DFNA5</i>	Executioner of pyroptosis	DFNB9	<i>OTOF</i>	Trafficking of membrane vesicles
DFNA6/14/38	<i>WFS1</i>	Transmembrane protein	DFNB12	<i>CDH23</i>	Intercellular adherence protein
DFNA7	<i>LMX1A</i>	Transcription factor	DFNB15/72/95	<i>GIPC3</i>	PDZ domain-containing protein
DFNA8/12	<i>TECTA</i>	Tectorial membrane protein	DFNB16	<i>STRC</i>	Stereocilia protein
DFNA9	<i>COCH</i>	Unknown	DFNB18	<i>USH1C</i>	Unknown
DFNA10	<i>EYA4</i>	Developmental gene	DFNB18B	<i>OTOG</i>	Tectorial membrane protein
DFNA11	<i>MYO7A</i>	Cytoskeletal protein	DFNB21	<i>TECTA</i>	Tectorial membrane protein
DFNA13	<i>COL11A2</i>	Cytoskeletal protein	DFNB22	<i>OTOA</i>	Gel attachment to nonsensory cell
DFNA15	<i>POU4F3</i>	Transcription factor	DFNB23	<i>PCDH15</i>	Morphogenesis and cohesion
DFNA17	<i>MYH9</i>	Cytoskeletal protein	DFNB24	<i>RDX</i>	Cytoskeletal protein
DFNA20/26	<i>ACTG1</i>	Cytoskeletal protein	DFNB25	<i>GRXCR1</i>	Reversible S-glutathionylation of proteins
DFNA22	<i>MYO6</i>	Unconventional myosin	DFNB26	<i>GAB1</i>	Member of insulin receptor substrate 1-like multisubstrate docking adapter protein family
DFNA23	<i>SIX1</i>	Developmental gene	DFNB28	<i>TRIOBP</i>	Cytoskeletal-organizing protein
DFNA25	<i>SLC17A8</i>	Vesicular glutamate transporter	DFNB29	<i>CLDN14</i>	Tight junctions
DFNA27	<i>REST</i>	Transcriptional repressor	DFNB30	<i>MYO3A</i>	Hybrid motor-signaling myosin
DFNA28	<i>GRHL2</i>	Transcription factor	DFNB31	<i>WHRN</i>	PDZ domain-containing protein
DFNA34	<i>NLRP3</i>	Pyrin-like protein involved in inflammation	DFNB32/105	<i>CDC14A</i>	Protein phosphatase involved in hair cell ciliogenesis
DFNA36	<i>TMC1</i>	Transmembrane protein	DFNB35	<i>ESRRB</i>	Estrogen-related receptor beta protein
DNA37	<i>COL11A1</i>	Cytoskeletal protein	DFNB36	<i>ESPN</i>	Ca-insensitive actin-bundling protein
DFNA40	<i>CRYM</i>	Thyroid hormone-binding protein	DFNB37	<i>MYO6</i>	Unconventional myosin
DFNA41	<i>P2RX2</i>	Purinergic receptor	DFNB39	<i>HFG</i>	Hepatocyte growth factor
DFNA44	<i>CCDC50</i>	Effector of epidermal growth factor-mediated signaling	DFNB42	<i>ILDR1</i>	Ig-like domain-containing receptor
DFNA50	<i>MIRN96</i>	MicroRNA	DFNB44	<i>ADCY1</i>	Adenylate cyclase
DFNA51	<i>TJP2</i>	Tight junction protein	DFNB48	<i>CIB2</i>	Calcium and integrin binding protein
DFNA56	<i>TNC</i>	Extracellular matrix protein	DFNB49	<i>BDP1</i>	Subunit of RNA polymerase
DFNA64	<i>SMAC/DIABLO</i>	Mitochondrial proapoptotic protein	DFNB49	<i>MARVELD2</i>	Tight junction protein
DFNA65	<i>TBC1D24</i>	ARF6-interacting protein	DFNB53	<i>COL11A2</i>	Collagen protein
DFNA66	<i>CD164</i>	Sialomucin	DFNB59	<i>PJVK</i>	Zn-binding protein
DFNA67	<i>OSBPL2</i>	Intracellular lipid receptor	DFNB60	<i>SLC22A4</i>	Prestin, motor protein of cochlear outer hair cell
DFNA68	<i>HOMER2</i>	Stereociliary scaffolding protein	DFNB61	<i>SLC26A5</i>	Motor protein
DFNA69	<i>KITLG</i>	Ligand for KIT receptor	DFNB63	<i>LRTOMT/COMT2</i>	Putative methyltransferase
DFNA70	<i>MCM2</i>	Initiation and elongation during DNA replication	DFNB66	<i>DCDC2</i>	Ciliary protein
DFNA73	<i>PTPRQ</i>	Member of type III receptor-like protein-tyrosine phosphatase (PTPase) family	DFNB66/67	<i>LHFPL5</i>	Tetraspan protein
	<i>DMXL2</i>	Regulator of Notch signaling	DFNB68	<i>S1PR2</i>	Tetraspan membrane protein of hair cell stereocilia
	<i>MYO3A</i>	Member of myosin superfamily	DFNB70	<i>PNPT1</i>	Mitochondrial-RNA-import protein
	<i>PDE1C</i>	Catalyze hydrolysis of cAMP and cGMP	DFNB73	<i>BSND</i>	Beta subunit of chloride channel
	<i>TRRAP</i>	Transformation/transcription domain associated protein	DFNB74	<i>MSRB3</i>	Methionine sulfoxide reductase
	<i>PLS1</i>	Actin-bundling protein	DFNB76	<i>SYNE4</i>	Part of LINC tethering complex
	<i>SCD5</i>	Catalyzes formation of monounsaturated fatty acids from saturated fatty acids	DFNB77	<i>LOXHD1</i>	Stereociliary protein
	<i>SLC12A2</i>	Sodium-potassium-chloride transporter	DFNB79	<i>TPRN</i>	Unknown
	<i>MAP1B</i>	Microtubule binding protein	DFNB82	<i>GPSM2</i>	G protein signaling modulator
	<i>RIPOR2/FAM65B</i>	Membrane-associated protein in stereocilia	DFNB84	<i>PTPRQ</i>	Type III receptor-like protein-tyrosine phosphatase family

(Continued)

TABLE 34-1 Hereditary Hearing Impairment Genes (Continued)

DESIGNATION	GENE	FUNCTION	DESIGNATION	GENE	FUNCTION
DFNB84	<i>OTOG</i>	Otogelin-like protein		<i>WBP2</i>	Transcriptional coactivator for estrogen receptor-alpha and progesterone receptor
DFNB86	<i>TBC1D24</i>	GTPase-activating protein		<i>ESRP1</i>	Modulates activation of G proteins
DFNB88	<i>ELMOD3</i>	GTPase-activating protein		<i>MPZL2</i>	Mediates epithelial cell-cell interactions in developing tissues
DFNB89	<i>KARS</i>	Lysyl-tRNA synthetase		<i>CEACAM16</i>	Cell adhesion molecule
DFNB91	<i>SERPINB6</i>	Protease inhibitor		<i>GRAP</i>	Cytoplasmic signaling protein
DFNB93	<i>CABP2</i>	Calcium-binding protein		<i>SPNS2</i>	Sphingosine-1-phosphate (S1P) transporter
DFN94	<i>NARS2</i>	Mitochondrial asparaginyl-tRNA synthetase		<i>CLDN9</i>	Tight junctions
DFNA97	<i>MET</i>	Oncogene/hepatocyte growth factor receptor		<i>CLRN2</i>	Maintenance of transducing stereocilia in auditory hair cells
DFNB98	<i>TSPEAR</i>	Epilepsy-associated repeats containing protein			
DFNB99	<i>TMEM132E</i>	Transmembrane protein			
DFNB100	<i>PPIP5K2</i>	Diphosphoinositol-pentakisphosphate kinase			
DFNB101	<i>GRXCR2</i>	Maintaining stereocilia bundles			
DFNB102	<i>EPS8</i>	Epidermal growth factor receptor			
DFNB103	<i>CLIC5</i>	Chloride ion transport			
DFNB104	<i>FAM65B/RIPOR2</i>	Membrane-associated protein in stereocilia			
DFNB106	<i>EPS8L2</i>	Actin remodeling in response to EGF stimulation			
DFNB108	<i>ROR1</i>	Receptor tyrosine kinase-like orphan receptor			
					X-linked
			DFNX1	<i>PRPS1</i>	Catalyzes phosphoribosylation of ribose 5-phosphate to 5-phosphoribosyl-1-pyrophosphate
			DFNX2	<i>POU3F4</i>	Transcription factor
			DFNX4	<i>SMPX</i>	Small muscle protein
			DFNX5	<i>AIFM1</i>	Mitochondrial flavin adenine dinucleotide (FAD)-dependent oxidoreductase
			DFNX6	<i>COL4A6</i>	Collagen protein

TABLE 34-2 Syndromic Hereditary Hearing Impairment Genes

SYNDROME	GENE	FUNCTION
Alport's syndrome	<i>COL4A3-5</i>	Cytoskeletal protein
BOR syndrome	<i>EYA1</i>	Developmental gene
	<i>SIX5</i>	Developmental gene
	<i>SIX1</i>	Developmental gene
Jervell and Lange-Nielsen syndrome	<i>KCNQ1</i>	Delayed rectifier K ⁺ channel
	<i>KCNE1</i>	Delayed rectifier K ⁺ channel
Norrie's disease	<i>NDP</i>	Cell–cell interactions
Pendred's syndrome	<i>SLC26A4</i>	Chloride/iodide transporter
	<i>FOXI1</i>	Transcriptional activator of <i>SLC26A4</i>
	<i>KCNJ10</i>	Inwardly rectifying K ⁺ channel
Treacher Collins syndrome	<i>TCOF1</i>	Nucleolar–cytoplasmic transport
	<i>POLR1D</i>	Subunit of RNA polymerases I and III
	<i>POLR1C</i>	Subunit of RNA polymerases I and III
Usher's syndrome	<i>MYO7A</i>	Cytoskeletal protein
	<i>USH1C</i>	Unknown
	<i>CDH23</i>	Intercellular adherence protein
	<i>PCDH15</i>	Cell adhesion molecule
	<i>SANS</i>	Harmonin-associated protein
	<i>CIB2</i>	Calcium- and integrin-binding protein
	<i>USH2A</i>	Cell adhesion molecule
	<i>VLGR1</i>	G protein-coupled receptor
	<i>WHRN</i>	PDZ domain-containing protein
	<i>CLRN1</i>	Cellular synapse protein
	<i>HARS</i>	Histidyl-tRNA synthetase
WS type I, III	<i>PDZD7</i>	PDZ domain-containing protein
	<i>PAX3</i>	Transcription factor
	<i>MITF</i>	Transcription factor
WS type II	<i>SNAI2</i>	Transcription factor
	<i>EDNRB</i>	Endothelin B receptor
WS type IV	<i>EDN3</i>	Endothelin B receptor ligand
	<i>SOX10</i>	Transcription factor

Abbreviations: BOR, branchio-oto-renal syndrome; WS, Waardenburg's syndrome.

acoustic-reflex decay. This test battery provides a screening evaluation of the entire auditory system and allows one to determine whether further differentiation of a sensory (cochlear) from a neural (retrocochlear) hearing loss is indicated.

Pure tone audiometry assesses hearing acuity for pure tones. The test is administered by an audiologist and is performed in a sound-attenuated chamber. The pure tone stimulus is delivered with an audiometer, an electronic device that allows the presentation of specific frequencies (generally between 250–8000 Hz) at specific intensities. Air- and bone-conduction thresholds are established for each ear. Air-conduction thresholds are determined by presenting the stimulus in air with the use of headphones. Bone-conduction thresholds are determined by placing the stem of a vibrating tuning fork or an oscillator of an audiometer in contact with the head. In the presence of a hearing loss, broad-spectrum noise is presented to the nontest ear for masking purposes so that responses are based on perception from the ear under test.

The responses are measured in decibels (dBs). An *audiogram* is a plot of intensity in dBs of hearing threshold versus frequency. A dB is equal to 20 times the logarithm of the ratio of the sound pressure required to achieve threshold in the patient to the sound pressure required to achieve threshold in a normal-hearing person. Therefore, a change of 6 dB represents doubling of sound pressure, and a change of 20 dB represents a tenfold change in sound pressure. Loudness, which depends on the frequency, intensity, and duration of a sound, doubles with approximately each 10-dB increase in sound pressure level. Pitch, on the other hand, does not directly correlate with frequency. The

TABLE 34-3 Signs and Symptoms Suggestive of Hearing Loss

Saying "huh" a great deal
Reduced clarity of hearing
Difficulty understanding conversations in background noise
Family complaining of hearing loss
Tinnitus
Turning the volume up on radio or television
Sensitivity to noises
Fullness in the ear
Avoiding social settings

perception of pitch changes slowly in the low and high frequencies. In the middle tones, which are important for human speech, pitch varies more rapidly with changes in frequency.

Pure tone audiometry establishes the presence and severity of hearing impairment, unilateral versus bilateral involvement, and the type of hearing loss. Conductive hearing losses with a large mass component, as is often seen in middle-ear effusions, produce elevation of thresholds that predominate in the higher frequencies. Conductive hearing losses with a large stiffness component, as in fixation of the footplate of the stapes in early otosclerosis, produce threshold elevations in the lower frequencies. Often, the conductive hearing loss involves all frequencies, suggesting involvement of both stiffness and mass. In general, sensorineural hearing losses such as presbycusis affect higher frequencies more than lower frequencies (Fig. 34-3). An exception is Ménière's disease, which is characteristically associated with low-frequency sensorineural hearing loss (though any frequency can be affected). Noise-induced hearing loss has an unusual pattern of hearing impairment in which the loss at 3000–4000 Hz is greater than at higher frequencies. Vestibular schwannomas characteristically affect the higher frequencies, but any pattern of hearing loss can be observed.

Speech recognition requires greater synchronous neural firing than is necessary for appreciation of pure tones. *Speech audiometry* tests the clarity with which one hears. The *speech reception threshold* (SRT) is defined as the intensity at which speech is recognized as a meaningful symbol and is obtained by presenting two-syllable words with an equal accent on each syllable. The intensity at which the patient can repeat 50% of the words correctly is the SRT. Once the SRT is determined, discrimination or word recognition ability is tested by presenting one-syllable words at 25–40 dB above the SRT. The words are phonetically balanced in that the phonemes (speech sounds) occur in the list of words at the same frequency that they occur in ordinary conversational English. An individual with normal hearing or conductive hearing loss can repeat 88–100% of the phonetically balanced words correctly. Patients with a sensorineural hearing loss have variable loss of discrimination. As a general rule, neural lesions produce greater deficits in discrimination than do cochlear lesions. For example, in a patient with mild asymmetric sensorineural hearing loss, a clue to the diagnosis of vestibular schwannoma is the presence of greater than expected deterioration in discrimination ability. Deterioration in discrimination ability at higher intensities above the SRT also suggests a lesion in the eighth nerve or central auditory pathways.

Tympanometry measures the impedance of the middle ear to sound and is useful in diagnosis of middle-ear effusions. A *tympanogram* is the graphic representation of change in impedance or compliance as the pressure in the ear canal is changed. Normally, the middle ear is most compliant at atmospheric pressure, and the compliance decreases as the pressure is increased or decreased (type A); this pattern is seen with normal hearing or in the presence of sensorineural hearing loss. Compliance that does not change with change in pressure suggests middle-ear effusion (type B). With a negative pressure in the middle ear, as with Eustachian tube obstruction, the point of maximal compliance occurs with negative pressure in the ear canal (type C). A tympanogram in which no point of maximal compliance can be obtained is most commonly seen with discontinuity of the ossicular chain (type A_d). A reduction in the maximal compliance peak can be seen in otosclerosis (type A_s).

During tympanometry, an intense tone elicits contraction of the stapedius muscle. The change in compliance of the middle ear with contraction of the stapedius muscle can be detected. The presence or absence of this *acoustic reflex* is important in determining the etiology of hearing loss as well as in the anatomic localization of facial nerve paralysis. The acoustic reflex can help differentiate between conductive hearing loss due to otosclerosis and that caused by an inner-ear "third window": it is absent in otosclerosis and present in inner-ear conductive hearing loss. Normal or elevated acoustic reflex thresholds in an individual with sensorineural hearing impairment suggest a cochlear hearing loss. An absent acoustic reflex in the setting of sensorineural hearing loss is not helpful in localizing the site of lesion. Assessment of *acoustic reflex decay* helps differentiate sensory from neural hearing losses. In neural hearing loss, such as with vestibular schwannoma, the reflex adapts or decays with time.

OAES generated by outer hair cells only can be measured with microphones inserted into the external auditory canal. The emissions may be spontaneous or evoked with sound stimulation. The presence of OAEs indicates that the outer hair cells of the organ of Corti are intact and can be used to assess auditory thresholds and to distinguish sensory from neural hearing losses.

Evoked Responses *Electrococleography* measures the earliest evoked potentials generated in the cochlea and the auditory nerve. Receptor potentials recorded include the cochlear microphonic, generated by the outer hair cells of the organ of Corti, and the summing potential, generated by the inner hair cells in response to sound. The whole nerve action potential representing the composite firing of the first-order neurons can also be recorded during electrococleography. Clinically, the test is useful in the diagnosis of Ménière's disease, in which an elevation of the ratio of summing potential to action potential is seen.

Brainstem auditory-evoked responses (BAERs), also known as ABRs, are useful in differentiating the site of sensorineural hearing loss. In response to sound, five distinct electrical potentials arising from different stations along the peripheral and central auditory pathway (eighth nerve, cochlear nucleus, superior olfactory complex, lateral lemniscus, and inferior colliculus) can be identified using computer averaging from scalp surface electrodes. BAERs are valuable in situations in which patients cannot or will not give reliable voluntary thresholds. They are also used to assess the integrity of the auditory nerve and brainstem in various clinical situations, including intraoperative monitoring, and in determination of brain death.

The *VEMP test* investigates otolith and vestibular nerve function by presenting a high-level acoustic stimulus and evoking a short-latency electromyographic potential; cVEMP (or cervical VEMP) and oVEMP (or ocular VEMP) have been described. The cVEMP elicits a vestibulocollic reflex whose afferent limb arises from acoustically sensitive cells in the saccule, with signals conducted via the inferior vestibular nerve. cVEMP is a biphasic, short-latency response recorded from the tonically contracted sternocleidomastoid muscle in response to loud auditory clicks or tones. cVEMPs may be diminished or absent in patients with early and late Ménière's disease, vestibular neuritis, benign paroxysmal positional vertigo, and vestibular schwannoma. On the other hand, the threshold for VEMPs may be lower in cases of superior canal dehiscence, other inner-ear dehiscence ("third window"), and perilymphatic fistula. The oVEMP, in contrast, is a response involving the utricle primarily and superior vestibular nerve. The oVEMP excitatory response is recorded from the extraocular muscle. The oVEMP is abnormal in superior vestibular neuritis.

Imaging Studies The choice of radiologic tests is largely determined by whether the goal is to evaluate the bony anatomy of the external, middle, and inner ear or to image the auditory nerve and brain. Axial and coronal CT of the temporal bone with fine 0.3-mm cuts is ideal for determining the caliber of the external auditory canal, integrity of the ossicular chain, and presence of middle-ear or mastoid disease; it can also detect inner-ear malformations. CT is also ideal for the detection of bone erosion with chronic otitis media and cholesteatoma. Pöschl reformatting in the plane of the superior semicircular canal is required for the identification of dehiscence or absence of bone over the superior semicircular canal. MRI is superior to CT for imaging of retrocochlear pathology such as vestibular schwannoma, meningioma, other lesions of the cerebellopontine angle, demyelinating lesions of the brainstem, and brain tumors. Both CT and MRI are equally capable of identifying inner-ear malformations and assessing cochlear patency for preoperative evaluation of patients for cochlear implantation.

TREATMENT

Disorders of the Sense of Hearing

In general, conductive hearing losses are amenable to surgical correction, whereas sensorineural hearing losses are usually managed medically. Atresia of the ear canal can be surgically repaired,

often with significant improvement in hearing. Alternatively, the conductive hearing loss associated with atresia can be addressed with a bone-anchored hearing aid (BAHA). Tympanic membrane perforations due to chronic otitis media or trauma can be repaired with an outpatient tympanoplasty. Likewise, conductive hearing loss associated with otosclerosis can be treated by stapedectomy, which is successful in >95% of cases. Tympanostomy tubes allow the prompt return of normal hearing in individuals with middle-ear effusions. Hearing aids are effective and well tolerated in patients with conductive hearing losses.

Patients with mild, moderate, and severe sensorineural hearing losses are regularly rehabilitated with hearing aids of varying configuration and strength. Hearing aids have been improved to provide greater fidelity and have been miniaturized. The current generation of hearing aids is nearly invisible, thus reducing stigma associated with their use. In general, the more severe the hearing impairment, the larger the hearing aid required for auditory rehabilitation. Digital hearing aids lend themselves to individual programming, and multiple and directional microphones at the ear level may be helpful in noisy surroundings. Because all hearing aids amplify noise as well as speech, the only absolute solution to the problem of noise is to place the microphone closer to the speaker than the noise source. This arrangement is not possible with a self-contained, cosmetically acceptable device. A significant limitation of rehabilitation with a hearing aid is that although it is able to enhance detection of sound with amplification, it cannot restore clarity of hearing that is lost with presbycusis.

The cost of a single hearing aid (~\$2300 US) is a significant obstacle for many hearing-impaired individuals and usually bilateral amplification is recommended. To reduce cost and spur innovation, a new category of over-the-counter amplification devices that can be purchased similar to reading eyeglasses by simply walking into a store has recently been approved by the US Food and Drug Administration. By reducing the cost of amplification devices to consumers, promoting innovation, and increasing competition, this new class of devices could fundamentally change the way hearing rehabilitation is delivered.

Patients with unilateral deafness have difficulty with sound localization and reduced clarity of hearing in background noise. They may benefit from a contralateral routing of signal (CROS) hearing aid in which a microphone is placed on the hearing-impaired side, and the sound is transmitted to the receiver placed on the contralateral ear. The same result may be obtained with a BAHA, in which a hearing aid clamps to a screw integrated into the skull on the hearing-impaired side. Like the CROS hearing aid, the BAHA transfers the acoustic signal to the contralateral hearing ear, but it does so by vibrating the skull. Patients with profound deafness on one side and some hearing loss in the better ear are candidates for a BICROS hearing aid; it differs from the CROS hearing aid in that the patient wears a hearing aid, and not simply a receiver, in the better ear. Unfortunately, while CROS and BAHA devices provide benefit, they do not restore hearing in the deaf ear. Only cochlear implants can restore hearing (see below). Increasingly, cochlear implants are being used for the treatment of patients with single-sided deafness; they show great promise in not only restoring hearing and reducing tinnitus, but also improving sound localization and performance in background noise.

In many situations, including lectures and the theater, hearing-impaired persons benefit from assistive devices that are based on the principle of having the speaker closer to the microphone than any source of noise. Assistive devices include infrared and frequency-modulated (FM) transmission as well as an electromagnetic loop around the room for transmission to the individual's hearing aid. Hearing aids with telecoils can also be used with properly equipped telephones in the same way. Bluetooth technology has revolutionized connectivity between hearing aids and other devices such as smart phones.

In the event that the hearing aid provides inadequate rehabilitation, cochlear implants may be appropriate (Fig. 34-4). Criteria for implantation include severe to profound hearing loss with open-set sentence cognition of ≤40% under best-aided conditions.

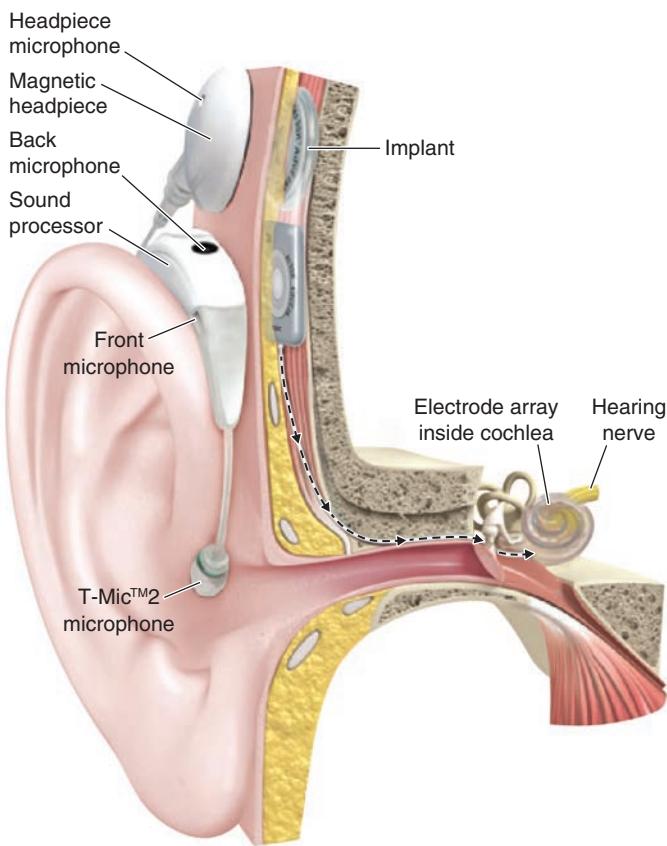


FIGURE 34-4 A cochlear implant is composed of an external microphone and speech processor worn on the ear and a receiver implanted underneath the temporalis muscle. The internal receiver is attached to an electrode that is placed surgically in the cochlea.

Worldwide, >600,000 hearing-impaired individuals have received cochlear implants. Cochlear implants are neural prostheses that convert sound energy to electrical energy and can be used to stimulate the auditory division of the eighth nerve directly. In most cases of profound hearing impairment, the auditory hair cells are lost but the ganglionic cells of the auditory division of the eighth nerve are preserved. Cochlear implants consist of electrodes that are inserted into the cochlea through the round window, speech processors that extract acoustic elements of speech for conversion to electrical currents, and a means of transmitting the electrical energy through the skin. Patients with implants experience sound that helps with speech reading, allows open-set word recognition, and helps in modulating the person's own voice. Usually, within the first 3–6 months after implantation, adult patients can understand speech without visual cues. With the current generation of multichannel cochlear implants, nearly 75% of patients are able to converse on the telephone. Bilateral cochlear implantations are commonly performed, especially in children; these patients perform better in background noise, have better sound localization, and are less fatigued by the "work" compared to monaural hearing.

Hybrid cochlear implants are indicated for the treatment of high-frequency hearing loss in patients who do not have profound hearing loss and yet do not benefit from hearing aids. Patients with presbycusis typically have normal low-frequency hearing while suffering from high-frequency hearing loss associated with loss of clarity that cannot always be adequately rehabilitated with a hearing aid. However, these patients are not candidates for conventional cochlear implants because they have too much residual hearing. The hybrid implant has been specifically designed for this patient population; it has a shorter electrode than a conventional cochlear implant and can be introduced into the cochlea atraumatically, thus preserving low-frequency hearing. Individuals with a hybrid implant use their own natural low-frequency "acoustic" hearing and

rely on the implant for providing “electrical” high-frequency hearing. Patients who have received the hybrid implant perform better on speech discrimination tests in both quiet and noisy backgrounds.

For individuals who were born without cochlea or have had both eighth nerves destroyed by trauma or bilateral vestibular schwannomas (e.g., neurofibromatosis type 2), brainstem auditory implants placed near the cochlear nucleus may provide auditory rehabilitation. Currently, brainstem implants provide sound awareness but unfortunately speech understanding remains elusive.

Tinnitus often accompanies hearing loss. Similar to background noise, tinnitus can degrade speech comprehension in individuals with hearing impairment. Patients with tinnitus should be advised to minimize caffeine ingestion, avoid high dosage of nonsteroidal anti-inflammatory drugs (NSAIDs), and reduce stress. Therapy for tinnitus is usually directed toward minimizing the appreciation of tinnitus. Relief of the tinnitus may be obtained by masking it with background music or white noise. Hearing aids are also helpful in tinnitus suppression, as are tinnitus maskers, devices that present a sound to the affected ear that is more pleasant to listen to than the tinnitus. The use of a tinnitus masker is often followed by several hours of inhibition of the tinnitus. Antidepressants have also been shown to be beneficial in helping patients cope with tinnitus.

Hard-of-hearing individuals often benefit from a reduction in unnecessary noise in the environment (e.g., radio or television) to enhance the signal-to-noise ratio. Speech comprehension is aided by lip reading; therefore, the impaired listener should be seated so that the face of the speaker is well illuminated and easily seen. Although speech should be in a loud, clear voice, one should be aware that in sensorineural hearing losses in general and in hard-of-hearing elderly in particular, recruitment (abnormal perception of loud sounds) may be troublesome. Above all, optimal communication cannot take place without both parties giving it their full and undivided attention.

■ PREVENTION

Conductive hearing losses may be prevented by prompt antibiotic therapy of adequate duration for AOM and by ventilation of the middle ear with tympanostomy tubes in middle-ear effusions lasting ≥ 12 weeks. Loss of vestibular function and deafness due to aminoglycoside antibiotics can largely be prevented by careful monitoring of serum peak and trough levels.

Some 10 million Americans have noise-induced hearing loss, and 20 million are exposed to hazardous noise in their employment. Noise-induced hearing loss can be prevented by avoidance of exposure to loud noise or by regular use of earplugs or fluid-filled ear muffs to attenuate intense sound. **Table 34-4** lists loudness levels for a variety of environmental sounds. High-risk activities for noise-induced hearing loss include use of electrical equipment for wood- and metalworking, and target practice or hunting with small firearms. All internal-combustion

TABLE 34-5 OSHA Daily Permissible Noise Level Exposure

SOUND LEVEL (dB)	DURATION PER DAY (h)
90	8
92	6
95	4
97	3
100	2
102	1.5
105	1
110	0.5
115	≤ 0.25

Note: Exposure to impulsive or impact noise should not exceed 140-dB peak sound pressure level.

Source: From https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=standards&p_id=9735.

and electric engines, including snow and leaf blowers, snowmobiles, outboard motors, and chainsaws, require protection of the user with hearing protectors. Virtually all noise-induced hearing loss is preventable through education, which should begin before the teenage years. Programs for conservation of hearing in the workplace are required by the Occupational Safety and Health Administration (OSHA) whenever the exposure over an 8-h period averages 85 dB. OSHA mandates that workers in such noisy environments have hearing monitoring and protection programs that include a preemployment screen, an annual audiologic assessment, and the mandatory use of hearing protectors. Exposure to loud sounds above 85 dB in the work environment is restricted by OSHA, with halving of allowed exposure time for each increment of 5 dB above this threshold; for example, exposure to 90 dB is permitted for 8 h; 95 dB for 4 h, and 100 dB for 2 h (**Table 34-5**).

■ FURTHER READING

- CARLSON ML: Cochlear implantation in adults. *N Engl J Med* 382:1531, 2020.
 ESPINOSA-SANCHEZ JM, LOPEZ-ESCAmez JA: Menière’s disease. *Handb Clin Neurol* 137:257, 2016.
 MOSER T, STARR A: Auditory neuropathy—neural and synaptic mechanisms. *Nat Rev Neurol* 12:135, 2016.
 PATEL M et al: Intratympanic methylprednisolone versus gentamicin in patients with unilateral Menière’s disease: a randomised, double-blind, comparative effectiveness trial. *Lancet* 388:2753, 2016.
 TIKKA C et al: Interventions to prevent occupational noise-induced hearing loss. *Cochrane Database Syst Rev* 7:CD006396, 2017.
 WILSON BS et al: Global hearing health care: new findings and perspectives. *Lancet* 390:2503, 2017.

TABLE 34-4 Decibel (Loudness) Level of Common Environmental Noise

SOURCE	DECIBEL (dB)
Weakest sound heard	0
Whisper	30
Normal conversation	55–65
City traffic inside car	85
OSHA Monitoring Requirement Begins	90
Jackhammer	95
Subway train at 200 ft	95
Power mower	107
Power saw	110
Painful Sound	125
Jet engine at 100 ft	140
12-gauge shotgun blast	165
Loudest sound that can occur	194

Abbreviation: OSHA, Occupational Safety and Health Administration.

35

Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat

Rachel L. Amdur, Jeffrey A. Linder



Upper respiratory symptoms are most commonly caused by viral infection but also can be caused by other infectious, inflammatory, allergic, autoimmune, and neoplastic conditions. This chapter will discuss ambulatory antibiotic prescribing and review the most common causes of upper respiratory symptoms, including nonspecific upper respiratory infections.

Ear pain is most commonly caused by otitis externa, acute otitis media (AOM), otitis media with effusion (OME), and acute mastoiditis. Sinus symptoms can be caused by acute sinusitis, invasive fungal sinusitis, nosocomial sinusitis, and chronic sinusitis. Sore throat and neck pain can be caused by streptococcal pharyngitis, nonstreptococcal pharyngitis, acute infectious mononucleosis, other types of bacterial pharyngitis, Lemierre's syndrome, gonococcal pharyngitis, diphtheria, acute HIV infection, head and neck abscesses, epiglottitis, and laryngitis. At the time of presentation, upper respiratory symptoms of most common viral and bacterial etiologies have generally lasted from hours up to a few days.

UPPER RESPIRATORY INFECTIONS

Upper respiratory infections (URIs) are acute respiratory infections that occur above the vocal cords. URIs, including nonspecific upper respiratory tract infection, otitis media, sinusitis, and pharyngitis, are collectively the most common symptomatic reason for seeking care in the United States. In terms of etiology, symptoms, and signs, URIs overlap with lower acute respiratory infections that occur below the vocal cords, such as influenza (Chap. 200), acute bronchitis, and pneumonia (Chap. 126), as well as with noninfectious cough (Chap. 38). The average adult has 2–4 URIs per year; children can have 6–10 URIs annually. URIs can be prevented by hand washing or sanitization, physical distancing, use of facial masks, isolation of persons who are ill, and environmental cleaning (Chap. 199).

SARS-CoV-2, the pathogen that causes COVID-19, can cause virtually any upper respiratory symptom (Chap. 199). COVID-19 symptoms appear 2–14 days after exposure and may include fever, chills, cough, shortness of breath, fatigue, myalgias, headaches, rhinorrhea, sore throat, nausea, vomiting, or diarrhea. New loss of taste or smell appears to be specific for COVID-19. Until there is widespread natural or vaccine-induced immunity, any respiratory symptom occurring in areas where SARS-CoV-2 is circulating should be considered a potential manifestation of COVID-19.

■ IMPROVING AMBULATORY ANTIBIOTIC PRESCRIBING

The only common acute respiratory infections that should be treated with antibiotics are AOM, sinusitis, streptococcal pharyngitis, and pneumonia. Even for AOM, sinusitis, and pharyngitis, only a minority of cases meet the criteria for antibiotic prescribing. Common respiratory viruses (Chap. 199) cause the overwhelming majority of acute respiratory infections, and these infections are generally self-limited; antibiotics neither speed resolution nor prevent complications for the majority of acute respiratory infections. Unfortunately, for this reason, at least half of ambulatory antibiotic prescriptions for acute respiratory infections in the United States are inappropriate. Internationally, population rates of antibiotic prescribing vary nearly threefold, with no differences in infectious complications. Antibiotics cause adverse drug effects, alter the microbiome, cause *Clostridioides difficile* infection (Chap. 134), increase health care costs, and increase the prevalence of antibiotic-resistant bacteria (Chap. 145).

Clinicians prescribe inappropriate antibiotics because of time pressure; fear of missing a rare bacterial diagnosis; concern about preventing a rare bacterial complication; a lack of salience of adverse antibiotic effects; or a mistaken belief that most patients expect, demand, or will not be satisfied without an antibiotic prescription.

■ AMBULATORY ANTIBIOTIC STEWARDSHIP

Antibiotic stewardship has traditionally been an inpatient concern (Chap. 144), but ambulatory antibiotic use accounts for ~85% of antibiotic use by patients in most developed countries. In 2016, the Centers for Disease Control and Prevention published the "Core Elements of Out-patient Antibiotic Stewardship." The core elements include (1) committing to improving antibiotic prescribing; (2) implementing at least one policy or practice to improve antibiotic prescribing and assessing its effectiveness; (3) monitoring antibiotic prescribing and providing feedback; and (4) providing educational resources to clinicians and

patients on antibiotic prescribing. Effective interventions to decrease inappropriate ambulatory antibiotic prescribing include peer comparison, accountable justification, precommitment, clinical decision support, patient education, and multifaceted interventions. Communication training has been particularly effective when it includes making a clear diagnosis, focusing on positive actions patients can take to feel better, reviewing the expected course of illness, and informing patients about concerning symptoms (red flags) for which they should seek or reconnect with care. Telemedicine—synchronous telephone or video or asynchronous electronic messaging—has the potential to improve patient convenience and reduce inappropriate antibiotic prescribing.

Several techniques that seemed promising for the reduction of ambulatory antibiotic prescribing remain unproven, have been ineffective (e.g., procalcitonin testing), or are not durable (e.g., C-reactive protein testing). The practice of delayed antibiotic prescription—i.e., a prescription given to a patient who is asked not to fill it unless symptoms do not improve in a few days—is conceptually flawed and should be avoided. Delayed antibiotic prescriptions are usually given for antibiotic-inappropriate diagnoses (e.g., viral infections); they ignore the natural history of acute respiratory infections, which are self-limited and generally last from 5 to 14 days; they put the burden of clinical decision-making on patients; and they send a confusing, mixed message to patients about the appropriateness of antibiotics for respiratory infections.

NONSPECIFIC UPPER RESPIRATORY INFECTION ("THE COMMON COLD")

■ DEFINITION AND ETIOLOGY

Nonspecific URI, or the common cold, is a respiratory tract infection in which no single symptom predominates. Nonspecific URI is most commonly caused by respiratory viruses that are acquired through direct contact with infected individuals, contaminated surfaces, and large and small respiratory droplets. The most common viral causes of nonspecific URIs are rhinoviruses (well over 100 serotypes; Chap. 199), coronaviruses, parainfluenza virus, respiratory syncytial virus, influenza virus (Chap. 199), adenovirus (57 serotypes; Chap. 199), metapneumovirus, and bocavirus (Chap. 199). Making a specific viral diagnosis is not practical, cost-effective, or necessary. Multiplex panels of reverse transcription polymerase chain reaction are available but may be overly sensitive, as prior recent infection can cause false-positive results. Although the diagnosis is usually obvious, clinicians diagnosing a nonspecific URI should also consider influenza (Chap. 200), measles (cough, coryza, and conjunctivitis; Chap. 205), acute HIV infection (in which sore throat and rash often predominate; see below and Chap. 202), and COVID-19 (Chap. 199).

Individual susceptibility to nonspecific URIs depends on prior exposure, immunity, general health, genetics, microbiome-related factors, and mental health and social factors, including stress. Prior exposure leads to immunity to specific rhinoviruses and adenoviruses, but the number of serotypes makes reinfection likely. Immunity to non-COVID-19 coronaviruses, parainfluenza virus, respiratory syncytial virus, and metapneumoviruses is generally weak or of short duration.

■ SYMPTOMS AND SIGNS

Common respiratory viruses have incubation periods of 2–8 days after exposure. Symptoms generally begin gradually and include nasal fullness or obstruction, rhinorrhea, sore throat, laryngitis, lymphadenopathy, cough, and low-grade fever. Patients may have myalgias, but this feature usually is not as prominent as it is in influenza. Epistaxis is common with frequent nose blowing.

On physical examination, findings vary, but patients may have conjunctivitis, pharyngeal erythema, pharyngeal exudates, or pharyngeal cobblestoning. Depending on the phase of illness, the nasal mucosa may be pale, boggy, or red and swollen. Nasal mucus can range from watery to purulent. On auscultation, the lungs may be clear, or the patient may have diffuse wheezing or bronchial breath sounds consistent with a viral infection. Symptoms usually last 5–10 days but often last up to 14 days.

Nonspecific Upper Respiratory Infection

For adults and older children, treatment of nonspecific URI is symptom-based. Fever, myalgias, and sore throat can be treated with acetaminophen or a nonsteroidal anti-inflammatory drug (NSAID) such as ibuprofen. Rhinorrhea can be treated with ipratropium bromide. Nasal congestion can be managed with nasal decongestants such as oxymetazoline (two sprays into each nostril twice a day for up to 5 days) or systemic decongestants such as pseudoephedrine. Products that combine a decongestant with analgesics, antihistamines, or both help relieve symptoms. Although supporting data are weak, cough may be relieved with dextromethorphan or benzonatate (Tessalon Perles). Opioids, while effective at relieving cough, are associated with somnolence, dysphoria, constipation, and addiction.

For children <6 years old, cough and cold medicines should not be prescribed, recommended, or used because of the risks of adverse effects. Honey can help soothe a sore throat for children >1 year old. Cool-mist humidifiers may help with breathing, and saline nasal drops and bulb suctioning can help with nasal congestion.

Patients need to be informed that symptoms generally peak early but can last for up to 14 days; that they are infectious as long as they have symptoms; and that they should rest and drink plenty of fluids to avoid dehydration. Red flags for which patients should seek care include a fever of >102°F, chest pain (other than from a pulled muscle), shortness of breath, dizziness, confusion, new ear or sinus pain, and symptoms lasting >14 days. Although nonspecific URI can be complicated by otitis media and bacterial sinusitis, for an individual patient, an antibiotic is more likely to cause an adverse reaction than to prevent complications.

Other remedies that are ineffective, of questionable benefit, or associated with significant adverse effects include echinacea, zinc, inhaled steam, vitamin C, vitamin D, garlic, antihistamines, Chinese medicinal herbs, intranasal glucocorticoids, *Pelargonium sidoides* herbal extract, saline nasal irrigation, and antiviral drugs.

EAR PAIN

Ear pain is most commonly caused by otitis externa and otitis media. In adults, otologic disease is almost always associated with hearing changes. At >50 years of age, temporal arteritis should be considered in patients who have headache, malaise, weight loss, fever, anorexia, and a normal ear exam. Head and neck cancers should be considered in persons with a history of smoking and alcohol use. In children, the presence of a foreign body should be considered.

Ear pain can also result from other causes of local infection, inflammation, trauma, or tumors or can be referred. Innervation of the ear and surrounding areas includes cranial nerves V, VII, IX, and X and cervical nerves C2 and C3. Neuropathic and myopathic pain syndromes (e.g., trigeminal neuralgia) can cause ear pain. Ramsay Hunt syndrome (herpes zoster oticus) (Chap. 441) and Bell's palsy (Chap. 441) are both associated with ear pain.

Dental pathology can cause pain that radiates to the ear; caries and abscesses are most common. Bruxism, malocclusion, and temporomandibular disorder may be associated with tenderness in muscular attachments and the temporomandibular joint. Salivary gland pathology and cervical adenopathy can cause pain that radiates to the ear.

Sinusitis, tonsillitis, and pharyngitis cause pain that can radiate to the ear via cranial nerve IX. Gastroesophageal reflux disease (Chap. 321) is often associated with ear symptoms. Myocardial infarction can cause ear pain via cranial nerve X.

Relapsing polychondritis (Chap. 366) is a rare condition associated with recurrent, sometimes bilateral, erythematous, or violaceous swelling of the auricle (sparring the earlobe). Inflammation from relapsing polychondritis can involve nasal septal, laryngeal, or respiratory cartilage and can cause ocular inflammation, audiovestibular damage, and nonerosive seronegative inflammatory arthritis.

■ OTITIS EXTERNA

Etiology and Clinical Manifestations Otitis externa is an inflammation or infection of the external auditory canal manifesting as pain, redness, swelling, aural discharge, and hearing impairment. It is often associated with bacterial infection (frequently by *Pseudomonas aeruginosa* or *Staphylococcus aureus*), but fungi like *Aspergillus* or *Candida* can be implicated.

Otitis externa is most common among preteen and teenage children. Risk factors for otitis externa include swimming (with the resulting condition referred to as "swimmer's ear," which is more common in the summer), mechanical trauma (from cotton swabs or hearing aids), narrow ear canals, cerumen obstruction, eczema, and psoriasis. Classic swimmer's ear is associated with bacterial infection. Physical exam is notable for pain on movement of the auricle or tragus and an external auditory canal that is erythematous, edematous, inflamed, and sometimes coated with exudate on otoscopy. In contrast, fungal otitis externa often manifests with pruritus and ear discharge but without much pain.

Otitis externa can co-occur with otitis media. Preauricular, mastoid, parotid, or cervical lymphadenopathy may be present. AOM with tympanic membrane rupture (see below) can be associated with ear discharge and debris in the ear canal but (unlike otitis externa) without sensitivity to movement of the auricle.

Malignant Otitis Externa Malignant otitis externa is a potentially life-threatening form of otitis externa that involves the temporal bone and occurs in patients with diabetes or other types of immunosuppression, often in older adults. Patients may have fever. Progression of malignant otitis externa can affect cranial nerve VII, IX, XI, or XII.

TREATMENT

Otitis Externa

Analgesia should be provided with acetaminophen or an NSAID. The mainstay of treatment is one or more topical antibacterial drugs with a glucocorticoid for 7–10 days. Polymyxin B-neomycin-hydrocortisone is often used but should be avoided in patients with tympanic membrane perforation because of ototoxicity. Ciprofloxacin-hydrocortisone is an alternative.

Topical aluminum acetate may be as effective as a topical antibacterial-glucocorticoid regimen. For patients whose condition does not improve within 2–4 days with topical treatment, ear wicks or gauze impregnated with or soaked in anti-infective agents can be placed. Ineffective treatments include oral antibiotics and topical antifungals. Otitis externa frequently recurs; its recurrence may be prevented with periodic acetic acid or aluminum acetate drops.

For malignant otitis externa, oral antipseudomonal antibiotics are often prescribed. Patients sometimes require IV pain medication, fluids, or other antimicrobials.

■ ACUTE OTITIS MEDIA

Epidemiology and Etiology AOM—for which patients almost always present within days—is predominantly a disease of children, with incidence peaking at 6–24 months of age. By age 6, ~60% of children will have had an episode of AOM. Younger children appear to be susceptible because of a shorter, more horizontal eustachian tube that more easily accumulates fluid than it does in older children and adults and because their immune system is still developing.

AOM is caused by a viral URI leading to edema and inflammation of the nasopharynx and eustachian tube, collection of fluid, and infection by bacteria that colonize the nasopharynx. Viruses isolated include respiratory syncytial virus, rhinoviruses, enteroviruses, coronaviruses, influenza virus, adenoviruses, and human metapneumovirus. The bacteria most commonly isolated are *Streptococcus pneumoniae*, non-typeable *Haemophilus influenzae*, and *Moraxella catarrhalis*.

Symptoms and Signs Symptoms of AOM include ear pain, fever, irritability, otorrhea, and anorexia. Physical examination may be notable for a bulging, inflamed, cloudy tympanic membrane, with obscured landmarks, and immobility of the membrane on pneumatoscopy, the Valsalva maneuver, or swallowing while holding the nose shut. (An immobile tympanic membrane is also indicative of perforation, old middle-ear adhesions, a blocked auditory tube, or the presence of middle-ear fluid.) Patients have conductive hearing loss. Severe signs and symptoms include moderate to severe otalgia, otalgia lasting at least 2 days, and a temperature of >102.2°F.

AOM should be diagnosed in children with moderate to severe bulging of the tympanic membrane or new-onset otorrhea (not due to otitis externa). With mild bulging of the tympanic membrane, AOM can also be diagnosed if the patient has had symptoms for <48 h or if there is intense erythema of the tympanic membrane. AOM should *not* be diagnosed in children who do not have middle-ear effusion.

TREATMENT

Acute Otitis Media

Pain from AOM should be treated with NSAIDs or acetaminophen, which are effective for mild to moderate pain. Topical agents like benzocaine, procaine, or lidocaine may provide some additional, brief benefit beyond that offered by NSAIDs or acetaminophen.

In up to 80% of children, AOM resolves without antibiotics. Indications for antibiotic treatment in children include an age of <6 months, bilateral ear findings in children 6 months to 2 years old, otorrhea in children >6 months old, and—in children of all ages—ear findings with severe otalgia, ear pain for >48 h, or a fever of >102.2°F (**Table 35-1**).

The benefits of antibiotics are modest and are offset by adverse effects. Antibiotics do not result in early resolution of pain but do decrease pain by day 2 or 3 (number needed to treat, 20 patients treated with antibiotics for 1 patient to have decreased pain by day 2 or 3). More children who receive antibiotics have vomiting, diarrhea, and rash (number needed to harm, 14 patients treated with antibiotics for 1 to have vomiting, diarrhea, or rash). Severe complications like mastoiditis are rare, and the number needed to treat to prevent a case of mastoiditis is ~5000 (i.e., 5000 otitis media patients treated with antibiotics to prevent 1 case of mastoiditis). The American Academy of Family Physicians recommends not routinely prescribing antibiotics for otitis media in children 2–12 years old who have nonsevere symptoms and for whom the observation option is reasonable.

The antibiotic of choice for AOM is high-dose amoxicillin (90 mg/kg per d, up to 3 g). Alternatives include cefdinir, cefuroxime, cefpodoxime, or IM ceftriaxone. If the patient has received amoxicillin in the prior 30 days, clinicians should prescribe amoxicillin/clavulanate (90/6.4 mg/kg per d) in two divided doses. The duration of antibiotic treatment is 10 days for children <2 years old or children with severe symptoms; 5–7 days for children 2–5 years old with mild to moderate AOM; and 5 days for children ≥6 years old with mild or moderate symptoms.

If a patient's condition is not better after 48–72 h of treatment, the antibiotic regimen should be changed to amoxicillin/clavulanate, a second- or third-generation oral cephalosporin, or IM ceftriaxone

for 3 days. If, despite a change in antibiotics, the patient's condition still does not improve, that patient should be referred to a specialist. Middle-ear effusions are present in 60–70% of children with AOM; these should resolve over 3 months. Tympanostomy tubes should be considered for recurrent AOM (i.e., three episodes in 6 months or four episodes in 1 year). Mastoiditis is a rare complication of AOM that is suggested by postauricular tenderness, a postauricular mass, or protrusion of the ear lobe.

In adults, AOM is rare and there is little high-quality evidence to guide treatment. For adults, it remains important to differentiate AOM from OME, but AOM is generally treated with antibiotics, regardless of bilaterality or otorrhea. Amoxicillin is the drug of choice. Adults should also be treated with decongestants and analgesics. Adults with more than two episodes in a year or persistent effusion should be referred to an otolaryngologist.

■ OTITIS MEDIA WITH EFFUSION

Definition and Etiology OME, also called serous otitis media, occurs when there is fluid in the middle ear but no acute infection. Most patients with OME are young children; >60% of cases occur in children <2 years old. Many children have recurrent episodes.

OME is most often a sequela of a viral infection causing AOM, but it can also be caused by allergies. In addition to allergies, predisposing factors include craniofacial abnormalities, gastroesophageal reflux, and enlarged adenoids.

Symptoms and Signs The most common symptoms are decreases in sound conduction and hearing. Children with OME may exhibit impaired language development or communication difficulties. More rarely, patients complain of intermittent ear fullness or earache, tinnitus, or balance problems. On examination, the tympanic membrane may be translucent or gray with fluid (often colorless or amber), air-fluid levels, or bubbles behind the membrane. There is a loss of the light reflex. The tympanic membrane has decreased mobility on pneumatic otoscopy. The evaluation may include audiology, tympanometry, and, in infants, measurement of auditory brainstem responses.

OME usually resolves spontaneously within 4–6 weeks. If it persists for >3 months, the condition is referred to as chronic OME or chronic serous otitis media.

Cholesteatomas are accumulations of epithelium or keratin in the middle ear that can enlarge, perforate the tympanic membrane, envelop the ossicles, or destroy surrounding tissue. Cholesteatomas can cause labyrinthitis, hearing loss, cranial nerve palsies, vertigo, meningitis, extradural or brain abscess, and lateral sinus thrombophlebitis.

TREATMENT

Otitis Media with Effusion

OME is treated with myringotomy with tympanostomy tube insertion. For young children with nasal obstruction or recurrent infection, adenoidectomy may be considered. Medications, including antihistamines, glucocorticoids, or antibiotics, do not reliably help. Children at risk for speech or language delay may need earlier referral for more aggressive treatment.

■ ACUTE MASTOIDITIS

Etiology Acute mastoiditis is a serious infection with significant morbidity despite antibiotic and surgical treatment. This condition is most common among children <2 years old but can occur at any age. Acute mastoiditis is often a complication of AOM but may develop without clinically apparent, prior AOM. In older children with acute mastoiditis, clinicians should suspect cholesteatoma.

The pathogenesis of mastoiditis involves spread of organisms from the middle-ear spaces through the aditus ad antrum to the mastoid air cells. *Incipient* mastoiditis consists of fluid within the mastoid air

TABLE 35-1 Indications for Antibiotic Treatment of Acute Otitis Media

AGE	INDICATION
<6 months	Antibiotic treatment reasonable for all
6 months to 2 years	Bilateral ear findings
≥6 months	Otorrhea
>2 years	Symptoms worsening or not improving within 48–72 h
All ages	Ear findings with severe otalgia, otalgia lasting at least 2 days, or temperature of >102.2°F

cells, without bony destruction of the bony septa, and can progress to *coalescent mastoiditis*, with destruction of the bony septa. Acute mastoiditis often causes subperiosteal abscess laterally. The organisms most commonly involved in mastoiditis are *S. pneumoniae*, *Streptococcus pyogenes*, *H. influenzae*, *S. aureus* (including methicillin-resistant *S. aureus* [MRSA] strains), and *P. aeruginosa*.

Symptoms and Signs Symptoms of acute mastoiditis include ear pain, fever, lethargy, or fussiness despite adequate treatment of AOM. Patients—especially those with subperiosteal abscess—may have postauricular erythema, tenderness, warmth, fluctuance, and protrusion of the auricle. Otoscopic examination most often yields findings of AOM and may show superoposterior protrusion of the external auditory canal. Complications of mastoiditis include facial nerve palsy, labyrinthitis, skull osteomyelitis, temporal lobe abscess, cerebellar abscess, meningitis, epidural abscess, subdural abscess, venous sinus thrombosis, or Bezold's abscess (an abscess medial to the sternocleidomastoid that tracks into the deep cervical fascia).

Evaluation Laboratory evaluation reveals elevation of inflammatory markers and white blood cells with neutrophilia. Imaging is not necessary in children with a classic history and presentation but may be required if there is concern about complications or severity. CT may show disruption of bony septations, fluid, mucosal thickening, periosteal thickening, disruption of the periosteum, or subperiosteal abscess. MRI with gadolinium permits better visualization of abscesses and vascular problems.

Differential Diagnosis The differential diagnosis of acute mastoiditis includes cellulitis, otitis externa, postauricular lymphadenopathy, perichondritis, and tumors, including rhabdomyosarcoma, Ewing sarcoma, and myofibroblastic tumor.

TREATMENT

Mastoiditis

Patients with mastoiditis should be admitted to the hospital and treated with IV antibiotics and myringotomy, with or without tympanostomy tubes; if there is no improvement within 48 h, mastoidectomy should be undertaken. Tympanostomy or myringotomy samples or subperiosteal abscess drainage should be sent for culture and sensitivity testing. Depending on complications, additional drainage and surgical procedures may be necessary.

Empirical IV antibiotic therapy for children without recurrent AOM or recent antibiotic treatment consists of vancomycin (if there is concern about antibiotic-resistant *S. pneumoniae* or MRSA) or a cephalosporin (e.g., cefepime or ceftazidime). Patients with recurrent AOM or recent antibiotic treatment should be given vancomycin plus an antipseudomonal penicillin. Culture and sensitivity results will guide antibiotic changes. IV antibiotic therapy should be continued for 7–10 days, and patients should complete a 4-week course of oral antibiotics.

SINUS SYMPTOMS

Sinus symptoms are commonly due to respiratory viruses. These symptoms are considered acute if they last <4 weeks, subacute if they last 4–12 weeks, and chronic if they last ≥12 weeks. Beyond sinus infection, the differential diagnosis of rhinitis includes the common cold, allergic rhinitis (Chap. 352), vasomotor rhinitis, rhinitis medicamentosa due to topical decongestants, drug-induced rhinitis (e.g., due to aspirin, ibuprofen, or beta blockers), autoimmune disease (e.g., granulomatosis with polyangiitis), and cerebrospinal fluid leak. Pain over the sinuses can be caused by headaches (Chap. 430), facial pain syndromes, temporomandibular disorder (Chap. 36), and dental pathology. Gastroesophageal reflux can cause referral of symptoms to the sinuses. Patients who have uncontrolled diabetes or are otherwise

immunocompromised can have rapidly progressing invasive fungal infections (Chap. 211). More indolent fungal infections should be considered in the event of recurrent or nonresolving sinusitis. In children, it is important to consider the presence of a foreign body as a cause of sinus symptoms.

ACUTE SINUSITIS

Definition and Etiology *Sinusitis* is an inflammation of the paranasal sinuses; *rhinosinusitis* also involves the nasal passages. The majority of acute sinusitis cases are caused by respiratory viruses. A diagnosis of sinusitis is a major reason for unnecessary antibiotic prescribing in adults: although <2% of sinusitis episodes are due to bacteria (most often *S. pneumoniae*, *H. influenzae*, or *M. catarrhalis*), antibiotics are prescribed at >70% of office visits for sinusitis. According to guideline criteria, no more than 50% of adults—and probably closer to 20%—meet the criteria for antibiotic prescribing.

Symptoms and Signs Sinusitis symptoms commonly include purulent nasal discharge, facial congestion or fullness, and facial pain or pressure. Other symptoms include fever; hyposmia or anosmia; ear pain, pressure, or fullness; postnasal drip; halitosis; maxillary toothache; cough; and fatigue. Risk factors for developing sinusitis include an age of 45–65 years, smoking, asthma, air travel, and allergies.

On physical examination, direct rhinoscopy reveals excess mucus or purulence. Patients may have tenderness over the maxillary sinuses and, in severe cases, erythema and swelling of the maxilla. Sinus transillumination is not accurate in diagnosing sinusitis.

Complications Complications from sinusitis can be dramatic but are extremely rare. These complications may include orbital cellulitis, osteomyelitis, meningitis, intracranial abscesses, and cavernous sinus thrombosis. New symptoms that might indicate a sinusitis complication include confusion, unilateral weakness, proptosis, limited ocular movements, and acute vision changes.

RECURRENT ACUTE SINUSITIS Patients who have four or more episodes of acute sinusitis in a year, without signs or symptoms between episodes, are said to have recurrent acute sinusitis.

INVASIVE FUNGAL SINUSITIS Invasive fungal sinusitis may develop in immunocompromised patients, such as those with uncontrolled diabetes or transplant recipients, and should be considered an emergency. Invasive fungal sinusitis is caused by *Mucorales* fungi or *Aspergillus* (Chap. 217). Patients may appear to have a rapidly progressive case of rhinosinusitis, with facial pain and pressure, headaches, and fever followed within days by cranial nerve involvement, orbital swelling, cellulitis, proptosis, chemosis, and ophthalmoplegia. Patients may be critically ill. Evaluation should include nasal endoscopy with biopsy and imaging with gadolinium-enhanced MRI as the preferred modality.

NOSOCOMIAL SINUSITIS Nosocomial sinusitis occurs in critically ill patients, often those who are nasotracheally intubated. Nosocomial sinusitis should be suspected in hospitalized patients who have fever without another identifiable cause.

TREATMENT

Acute Sinusitis

All patients with acute sinusitis should be counseled about symptom-based treatments, which may include decongestants, analgesic/antipyretics, nasal saline, or intranasal glucocorticoids. Intranasal decongestants (e.g., oxymetazoline, two sprays in each nostril twice a day for no more than 5 days) and oral decongestants (e.g., 12-h pseudoephedrine [120 mg] during the day) relieve pain, pressure, and rhinorrhea. Analgesics and antipyretics like acetaminophen or NSAIDs (e.g., ibuprofen), nasal saline spray, and nasal washes provide relief. Intranasal glucocorticoids may help, particularly for

TABLE 35-2 Indications for Antibiotic Treatment of Acute Sinusitis

INDICATION	DEFINITION
Persistent	Symptoms lasting ≥10 days
Severe	Fever of >102°F and either purulent nasal discharge or nasal pain for at least 3–4 consecutive days
Worsening	New fever, headache, or increase in nasal discharge following an upper respiratory tract infection that lasted for 5–6 days and was initially improving

Note: In typical populations, roughly 20% and no more than 50% of adults with sinusitis will meet the criteria for antibiotic prescribing.

patients with an allergic cause of sinusitis. Because patients may be accustomed to receiving antibiotics, provision of a clear explanation, symptom-based treatments, and reasons for reconsultation are important. Red flags for which patients should reconsult include recurrent fever of >102°F, sinus symptoms that worsen after initial improvement, and rapid worsening of facial pain that becomes persistent, as well as any other concerning symptoms.

Antibiotic prescribing criteria for sinusitis are based on symptoms (**Table 35-2**). Only patients with persistent, severe, or worsening symptoms, especially those who have already used decongestants and analgesics for 2–4 days, meet the criteria for antibiotic prescribing. The antibiotic of choice is amoxicillin/clavulanate (875 mg/125 mg bid for 7 days). Amoxicillin (875 mg PO bid for 7 days) is an alternative. For patients with mild penicillin allergies, cefuroxime is a reasonable choice. For those with severe penicillin allergies, doxycycline is a reasonable alternative. Macrolides are specifically not recommended for sinusitis because of high rates of macrolide-resistant *S. pneumoniae*.

Patients who meet the criteria for antibiotic prescribing should show signs of improvement after 3–5 days of therapy. If not, second-line regimens include amoxicillin/clavulanate (2000 mg/125 mg bid for 7 days) or levofloxacin, although fluoroquinolones are associated with dysglycemia, neuropathy, and tendon and aortic rupture. For patients whose condition still is not improving after 3–5 days of treatment with a second-line antibiotic or in whom a complication or an alternative diagnosis is suspected, clinicians should consider referral to an otorhinolaryngologist and/or the performance of imaging tests. The imaging modality of choice is noncontrast CT. Patients with recurrent acute sinusitis may benefit from nasal culture during episodes; imaging between episodes to identify predisposing anatomic abnormalities; and allergic or immunologic evaluation.

Patients with acute fungal sinusitis should be treated with IV antifungal agents and often require surgical debridement. Patients with nosocomial sinusitis should have precipitating factors (e.g., nasotracheal intubation) addressed and should be empirically treated with broad-spectrum antibiotics until culture and susceptibility results are available.

CHRONIC SINUSITIS

Definition and Etiology Chronic sinusitis is defined as inflammation of the paranasal sinuses that lasts >12 weeks. Chronic sinusitis is primarily an inflammatory disease and can also be associated with acute or chronic infection or allergic, structural (e.g., deviated nasal septum or polyps), and immunologic etiologies. Repeated viral infections may lead to chronic sinusitis. Bacterial colonization or chronic infection plays a role in some cases of chronic sinusitis. *S. aureus* and gram-negative bacteria are commonly identified. Commonly involved allergens and irritants are dust mites, mold, tobacco smoke, occupational factors, and other airborne toxins. Functional or immunologic problems can include impaired mucociliary clearance (e.g., due to cystic fibrosis) or immunodeficiency due to acquired conditions or medications. Chronic sinusitis often coexists with allergic rhinitis and asthma.

Symptoms and Signs Cardinal symptoms of chronic sinusitis are facial pain or pressure, nasal discharge or postnasal drip, congestion, and hyposmia or anosmia. Associated symptoms may include fatigue, malaise, ear pressure, hoarseness, and cough. The diagnosis of sinus inflammation must be confirmed with anterior rhinoscopy, nasal endoscopy, or imaging because up to 40% of patients with chronic sinus symptoms do not have mucosal changes evidencing disease.

In practical terms, chronic sinusitis can be divided into three main types (in decreasing order of frequency): (1) chronic sinusitis without polyps, (2) chronic sinusitis with polyps, and (3) allergic fungal sinusitis. In general, chronic sinusitis without polyps is more common among women, develops in childhood and young adulthood, is characterized by presentations with facial pain, and is often due to T_H1 lymphocyte predominance associated with bacterial infection or colonization. Chronic sinusitis with polyps is more common among men; develops in adulthood; is characterized by presentations with decrease or loss of smell, asthma, or aspirin sensitivity (**Chap. 287**); and is often due to T_H2 lymphocyte predominance associated with eosinophilic inflammation, asthma, or aspirin sensitivity. Allergic fungal rhinosinusitis is also associated with polyp formation; typically occurs in patients in their 20s and 30s who are from warm, humid regions and who have other atopic diseases; and is associated with IgE-mediated allergy and eosinophils (**Chap. 217**). The mucus in allergic fungal rhinosinusitis is classically greenish-brown, has a peanut butter-like consistency, and includes viable hyphae from *Aspergillus* or other fungal species. Allergic fungal rhinosinusitis is resistant to medical treatments.

Evaluation On anterior rhinoscopy, polyps are seen as white, gray, tan, or yellow translucent growths in the middle meatus. The imaging modality of choice is noncontrast CT. Allergic fungal rhinosinusitis may be unilateral; however, unilateral symptoms or polyps on exam or imaging, especially if associated with bloody discharge, should raise concern about tumors.

TREATMENT

Chronic Sinusitis

Treatment includes avoidance of identifiable triggers such as allergens, smoke, and irritants. Saline sprays and washes provide symptom relief, and higher-volume saline washes are probably more effective. Intranasal glucocorticoids, including mometasone and fluticasone sprays or higher-potency and higher-volume budesonide rinses, are mainstays of treatment, especially for chronic sinusitis with polyps. Intranasal glucocorticoids reduce polyp size. Oral administration of glucocorticoids for 2–3 weeks is sometimes effective against chronic sinusitis that is unresponsive to intranasal steroids—again, especially for patients with polyps. Intranasal or systemic antihistamines may help patients whose illness has an allergic component. Likewise, leukotriene antagonists like montelukast may help.

Although antibiotics are frequently prescribed for 2–4 weeks to patients with chronic sinusitis, there is little evidence that these drugs are effective. Evidence of modest quality supports the use of 3 months of macrolide treatment for patients who have chronic sinusitis without polyps. Antifungal agents have not shown benefit against any subtype of chronic sinusitis. Decongestants should be used only sparingly and briefly.

Endoscopic sinus surgery improves quality of life in patients who have had inadequate responses to medical therapy. Patients with more limited, focal disease may more reliably have better results. The goals of surgery are to remove polyps from the nasal cavity and paranasal sinuses. For patients with allergic fungal rhinosinusitis, medical therapy is classically ineffective, surgery produces good results, and patients should be treated with perioperative glucocorticoids. In children, adenoidectomy may be effective in some cases. In the future, immune endotyping may allow selection of more individualized biological treatments.

TABLE 35-3 Clinical Findings That Suggest Various Forms of Nonstreptococcal Pharyngitis

CLINICAL FINDING(S) OR BEHAVIORAL FACTOR	SUSPECTED DIAGNOSIS
Scarlatiniform rash	Group A β -hemolytic streptococci or <i>Arcanobacterium haemolyticum</i>
Cough and otitis media	<i>Haemophilus influenzae</i>
Sex between men with associated urogenital symptoms, fellatio between a woman and a man who has current urogenital symptoms, persistent sore throat unresponsive to penicillin	<i>Neisseria gonorrhoeae</i>
Travel to endemic areas, pseudomembrane on examination	<i>Corynebacterium diphtheriae</i>
Persistent sore throat with bronchopulmonary symptoms	<i>Mycoplasma pneumoniae</i>
Marked adenopathy (especially that involving posterior cervical or auricular nodes), splenomegaly, palatine petechiae, gelatinous uvula	Acute infectious mononucleosis
New sexual partner in the previous month; fever, rash, myalgias, headache	Acute HIV infection

SORE THROAT AND NECK PAIN

Sore throat is not synonymous with pharyngitis and can also be caused by submandibular space, retropharyngeal and peritonsillar abscesses, thyroiditis, gastroesophageal reflux, tumors, and postnasal drainage.

Acute pharyngitis, in which symptoms are generally present for days, is most often caused by respiratory viruses; is often caused by group A β -hemolytic streptococci (GAS); and can be caused by other bacteria (including *Neisseria gonorrhoeae*), Epstein-Barr virus (EBV), and HIV. On physical examination, pharyngeal erythema is associated most commonly with viral infections, including the common cold and influenza. Pharyngeal exudate should not be confused with *Candida* infection, which looks like cottage-cheese, can be scraped off, and leaves a bleeding surface, or leukoplakia, which cannot be scraped off. History and exam findings may help differentiate sore throat and pharyngitis of various etiologies (Table 35-3).

■ STREPTOCOCCAL PHARYNGITIS

GAS is the only common cause of sore throat that should be treated with antibiotics. The principal goal in the evaluation of adults with sore throat is to identify patients likely to have GAS pharyngitis, or “strep throat.” Prompt antibiotic treatment of adults likely to have strep throat has the potential to reduce symptoms, prevent the spread of disease, and reduce suppurative complications (e.g., peritonsillar abscess). Nonsuppurative complications are rare. In developed countries, the prevalence of rheumatic fever (Chap. 148) is extremely low, and antibiotic treatment does not prevent poststreptococcal glomerulonephritis (Chap. 148).

Most patients with non-GAS pharyngitis have various forms of viral pharyngitis and do not require antibiotics. Nevertheless, clinicians prescribe antibiotics to a majority of adults with sore throats. By using a simple clinical scoring algorithm, clinicians can predict the presence or absence of GAS with sufficient accuracy and avoid prescribing antibiotics to patients who are unlikely to have strep throat. Although there is a role for testing (see “Evaluation,” below), most adults with sore throat do not need to have a GAS test.

About 10% of adults with sore throat are infected with GAS. Among children with sore throat, the prevalence of GAS can be as high as 35%, with rates peaking from 5 to 15 years of age. The prevalence of GAS is higher in winter and early spring. The risk of streptococcal pharyngitis is elevated among health care and child care workers, teachers, parents of young children, and patients exposed to individuals with strep throat. Clinicians need to be aware of local outbreaks of GAS infection,

TABLE 35-4 The Centor Criteria and the Probability of Streptococcal Pharyngitis for Adults^a

NO. OF CRITERIA MET ^b	POSTEVALUATION PROBABILITY (%)	RECOMMENDATION
0	2	No test, no antibiotic
1	3	No test, no antibiotic
2	8	Rapid test
3	19	Rapid test
4	41	Empirical antibiotic treatment or rapid test

^aAssuming a pretest probability of strep throat for adults of 10%. ^bThe criteria are (1) a history of fever, (2) an absence of cough, (3) tender anterior cervical lymphadenopathy, and (4) tonsillar swelling or exudate. Each criterion gets 1 point. Roughly 40–60% of adults will meet no criteria or one criterion; ~20% will meet the criteria for antibiotic prescribing.

particularly in military and institutional settings, where the prevalence of GAS and the risk of acute rheumatic fever may be elevated.

Evaluation The Centor criteria consist of four findings, each of which is assigned 1 point: (1) history of fever, (2) absence of cough, (3) tender anterior cervical lymphadenopathy, and (4) tonsillar exudate or swelling. The Centor criteria are easy to assess and accurately stratify adult patients with suspected streptococcal pharyngitis. Patients with no points have a 2% probability of being infected with GAS, whereas those with 4 points have a probability of 41% (Table 35-4). The Centor criteria have an area under the curve of 0.79. Other clinical decision algorithms similar to the Centor criteria may not perform as well, are not as simple, or have not been as rigorously evaluated.

If the test/no treatment threshold is set at 5%, for a GAS prevalence of ~10%, adults meeting no criteria or only one Centor criterion have a probability of GAS pharyngitis so low that they should neither be tested nor be treated with an antibiotic. Adults meeting two or three Centor criteria have an intermediate probability of GAS pharyngitis; they should have a rapid antigen test performed, and the results should guide antibiotic treatment. For adults meeting four Centor criteria, it is reasonable either to perform a rapid test or to institute empirical antibiotic treatment. However, some guidelines recommend—and some ambulatory quality measures require—a GAS test to be associated with antibiotic prescribing in adults, regardless of the number of Centor criteria met.

In children, the Centor criteria are less specific, and streptococcal pharyngitis should be confirmed with testing. Children who have signs of pharyngitis without signs of viral infection (conjunctivitis, runny nose, cough, hoarseness, nonexudative oral lesions) should have testing performed.

Outside of the United States, because complications are rare and even streptococcal pharyngitis is self-limited in the vast majority of cases, some guidelines do not recommend use of rapid GAS testing or routine antibiotic treatment of sore throat.

Clinicians should have a lower threshold for diagnosing and treating GAS pharyngitis in patients with a history of acute rheumatic fever, patients with documented streptococcal exposure in the past week, patients who live in a community with a current strep throat epidemic, and patients who are diabetic or otherwise immunocompromised.

RAPID STREP TESTS Rapid GAS-specific antigen tests have a sensitivity of ~80% and a specificity of ~95%. Results are available within minutes and can be used to make therapeutic decisions before the patient leaves the office. Improper collection technique can adversely affect the sensitivity of rapid strep tests: clinicians should rub the tonsils and pharynx, touching any areas where exudate or ulceration are present.

THROAT CULTURES A single-swab throat culture has a sensitivity of ~85–90%, as defined by isolation of GAS on a second swab. A throat culture can also be falsely positive for true infection: some patients with a culture positive for GAS may be only uninfected carriers, as defined by their failure to exhibit a fourfold increase in antibodies to GAS—the gold standard test. Among adults and children seeking medical care for a sore throat, test specificity may be as low as 50–70% because of

patients who do not exhibit serologic evidence of infection. Throat cultures are not recommended for the routine evaluation of adults with sore throat. The modest gain in sensitivity over rapid testing is outweighed by the 24- to 48-h delay in test results, with a consequent delay in the symptomatic relief associated with antibiotic treatment.

Indiscriminate strep testing in adults with sore throat or respiratory symptoms should be discouraged. Rapid strep tests and culture do not differentiate between patients who have true infection and those who are carriers of GAS (with carriage rates as high as 20% among schoolchildren and ~5% among adolescents and young adults). In adults who meet no Centor criteria or only one criterion—40–60% of adults with pharyngitis—a positive test is highly likely to be falsely positive and/or to represent GAS carriage.

Complications Complications of streptococcal pharyngitis are rare but include acute rheumatic fever (Chap. 148), poststreptococcal glomerulonephritis (Chap. 148), scarlet fever (Chap. 148), sinusitis, peritonsillar abscess, and other invasive GAS infections.

TREATMENT

Streptococcal Pharyngitis

All patients with pharyngitis—nonstreptococcal and streptococcal—should receive analgesics (acetaminophen or NSAIDs). Saline gargles, humidification, soft foods, and tea with honey soothe a painful throat.

Penicillin is the antibiotic of choice for streptococcal pharyngitis (Table 35-5). Penicillin is a narrow-spectrum, low-cost, and well-tolerated drug to which no GAS isolate has been resistant. Amoxicillin is an acceptable alternative in children as it comes in a palatable liquid form. For patients with mild penicillin allergy, cephalaxin and cefadroxil are good alternatives. For patients with severe penicillin allergies, clinicians should prescribe erythromycin, clarithromycin, or clindamycin. Unlike other infections for which emerging evidence supports progressively shorter antibiotic courses, streptococcal pharyngitis requires longer courses (7–10 days), which are more effective.

Glucocorticoids (e.g., dexamethasone, 10 mg as a single oral dose) have so far been poorly studied as an adjunctive treatment for sore throat and strep throat and are not recommended. These drugs may result in decreased pain within 24 h but do not decrease school or work absenteeism or relapse rates. Even short courses of steroids are associated with increased rates of sepsis, gastrointestinal bleeding, congestive heart failure, venous thromboembolism, and fracture within 30 days.

Streptococcal and nonstreptococcal pharyngitis should resolve in 3–5 days. Symptoms that should lead patients to seek further care include shaking chills (rigors), neck swelling (beyond lymphadenopathy), trouble swallowing, drooling, or symptoms that persist for >5 days without improvement.

TABLE 35-5 Antibiotic Treatment of Group A Streptococcal Pharyngitis

ANTIBIOTIC	DOSING
Antibiotic of Choice	
Penicillin	500 mg PO qid or 1000 mg PO bid × 10 days
Alternative for Non-Penicillin-Allergic Patients	
Amoxicillin	500 mg PO bid or 1000 mg qd × 10 days
Alternatives for Non-Anaphylactic Penicillin-Allergic Patients	
Cephalaxin	500 mg PO bid × 10 days
Cefadroxil	1 g PO qd × 10 days
Alternatives for Patients with Severe Penicillin Allergy	
Erythromycin	250–500 mg PO qid or 500–1000 mg PO bid × 5 days
Clarithromycin	500 mg PO bid × 5 days
Clindamycin	300 mg PO tid × 10 days

NONSTREPTOCOCCAL PHARYNGITIS

Acute Infectious Mononucleosis New EBV infection may be the cause of pharyngitis in 1–6% of young adults (Chap. 194). EBV is rarely the cause of pharyngitis in adults >40 years of age. The full-blown acute syndrome, which is present in only about one-fourth of patients with infectious mononucleosis (“mono”), is characterized by a triad of clinical, hematologic, and serologic findings. The clinical presentation is typified by the development over several days of malaise, fever, sore throat, and marked adenopathy that is particularly evident in the cervical lymph nodes. On physical examination, marked adenopathy is virtually always documented and is most specific for mononucleosis when the posterior cervical or posterior auricular nodes are involved. Splenomegaly and exudative pharyngitis with prominent tonsillar swelling, palatine petechiae, and a gelatinous uvula are often noted. The classic hematologic findings are an absolute lymphocyte count of >4000/ μ L or a relative lymphocyte count of >50% with “atypical” morphologic features in >10% of the lymphocytes. The characteristic serologic finding is the heterophil antibody, which is detectable in only 40% of patients during the first week of illness but in 80–90% of patients by the third week.

Other Bacterial Pharyngitis Non-group A streptococci (especially group C and group G streptococci), *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, *N. gonorrhoeae*, and *H. influenzae* have all been associated with sore throat in some studies. Although antibacterial treatment has not been proven to speed the resolution of symptoms and signs of any of these types of nonstreptococcal pharyngitis, antibiotic treatment is indicated if throat cultures from a patient with persistent sore throat yield group C or group G streptococci.

LEMIERRE'S SYNDROME Lemierre's syndrome consists of septic thrombophlebitis of the internal jugular vein accompanied by metastatic infections, most commonly of the lung but with possible involvement of the joints, bones, liver, meninges, and brain. Lemierre's syndrome is most commonly caused by *Fusobacterium necrophorum*, although it can also be caused by species of *Bacteroides*, *Eikenella*, *Streptococcus*, *Peptostreptococcus*, or other bacterial genera. This syndrome probably occurs predominantly in male patients. Clinicians should consider Lemierre's syndrome in a teenage or young adult patient who has non-GAS pharyngitis that is not resolving, particularly if it is accompanied by rigors, neck pain or swelling, or other extrapharyngeal symptoms.

GONOCOCCAL PHARYNGITIS *N. gonorrhoeae* may be the cause of pharyngitis in 1% of adult patients seeking primary care for a sore throat, although gonococcal infection of the pharynx is more often asymptomatic. When symptomatic, pharyngeal gonorrhea may range from mild to severe, with protracted pharyngitis characterized by pain, fever, and pharyngeal exudate. Gonococcal pharyngitis should be suspected in men who have sex with men with associated symptoms of urogenital infection, women who have practiced fellatio with a man with genital gonorrhea, and anyone who has persistent sore throat that has been unresponsive to treatment for presumptive streptococcal pharyngitis.

DIPHTHERIA Diphtheria, caused by *Corynebacterium diphtheriae*, is endemic in developing countries (Chap. 150). Diphtheria produces only mild pharyngitis beneath its characteristic grayish pseudomembrane.

ACUTE HIV INFECTION Clinicians should consider acute HIV infection in patients with sore throat, particularly when it is associated with headache, fever, myalgias, lymphadenopathy, anorexia, and rash (Chap. 202). Of patients with acute HIV infection, roughly half have a sore throat. However, in most settings in the United States, only ~1% of patients with viral or mononucleosis-like symptoms have acute HIV infection.

HEAD AND NECK ABSCESESSES

Head and neck abscesses are more common among patients with diabetes, who are immunocompromised, and among older adults. Such abscesses are often a complication of infections of the teeth and gums, throat,

or salivary ducts; lymphadenitis; ear infections; sinus infections; congenital cysts; and IV drug use. Prompt recognition is important, as head and neck abscesses can cause airway compromise due to edema or mass effect. Head and neck abscesses can follow fascial planes and spread to the mediastinum (where they can cause mediastinitis, pleural effusions, empyema, or pericarditis), the carotid sheath, the skull base, and the meninges. Head and neck abscesses have also been associated with aspiration pneumonia, necrotizing fasciitis, Lemierre's syndrome, and toxic shock syndrome.

Submandibular abscesses generally result from an infected or extracted tooth and can cause Ludwig angina, a swelling of the floor of the mouth that can enlarge and displace the tongue posteriorly.

Peritonsillar abscesses, which may occur predominantly in male patients, generally result from complicated bacterial pharyngitis and present with fever, dysphagia, profound throat pain (necessitating drooling to avoid swallowing saliva), trismus, and "hot potato voice" (inability to articulate, as if patients have hot food in their mouths). Patients are likely to have unilateral palate bulging, often with uvular deviation. Peritonsillar abscesses are caused by viridans group streptococci, β-hemolytic streptococci, *F. necrophorum*, *S. aureus*, *Prevotella*, and *Bacteroides*.

Retropharyngeal abscesses often present after an antecedent URI in children with sore throat, dysphagia, deep neck pain, neck stiffness, trismus, and drooling. The pharyngeal wall may be displaced, but swelling or abscess may not be apparent on examination. In severe cases, patients may have dyspnea and stridor.

Patients with suspected head and neck abscesses, with the possible exception of patients who have obvious peritonsillar abscesses, should undergo imaging by CT.

TREATMENT

Head and Neck Abscesses

The mainstays of treatment for head and neck abscesses are securing the airway, surgical drainage, and IV antibiotic administration. To secure the airway, mask ventilation or oral intubation may not be effective, and oral fiberoptic intubation or tracheotomy may be necessary. Peritonsillar abscess may be managed with needle aspiration and/or tonsillectomy. Other head and neck abscesses require incision and drainage. The selected IV antibiotics should cover streptococci, anaerobes, and possibly *S. aureus*. Frequently used antibiotics include ampicillin/sulbactam, clindamycin plus ceftriaxone, or meropenem. For some abscesses with adequate source control with incision and drainage, penicillin may be as effective as broader-spectrum agents.

■ EPIGLOTTITIS

Along with associated dysphagia, odynophagia, hoarseness, and stridor or tachypnea, supraglottitis or epiglottitis must be considered in adults presenting with sore throat. The inflamed and enlarged epiglottis protrudes up into the oropharynx. Patients may extend their neck or lean forward and drool oral secretions to avoid swallowing. Epiglottitis can cause "hot potato voice." Attempts to examine or swab the posterior pharynx or obtain a culture can provoke laryngospasm and should only be done carefully in a controlled setting. Because obstruction of the airway may become acutely life-threatening, the patient with epiglottitis must be observed in a hospital setting, and examination in an operating room, where an airway can be established immediately by an experienced operator, should be strongly considered. Although not necessary for the diagnosis, a lateral neck radiograph can demonstrate epiglottal swelling referred to as the "thumb sign."

In adults, conservative therapy under observation is sufficient in most cases, but intubation by an experienced clinician or tracheostomy may become necessary. Treatments also include humidification with nebulized normal saline or humidified oxygen and administration of glucocorticoids, IV antibiotics, and nebulized epinephrine.

H. influenzae, the most common cause of supraglottitis in children, is less common in adults. Other responsible organisms in adults are *S. pneumoniae*, *S. pyogenes*, and *S. aureus*. The *H. influenzae* type b vaccine has led to a dramatic decrease in epiglottitis overall, with large reductions in young children; however, the incidence of supraglottitis and epiglottitis in adults may be increasing.

■ LARYNGITIS

Laryngitis—*inflammation of the larynx and surrounding structures*—is most commonly caused by viral URIs. In children, parainfluenza virus can cause croup, or laryngotracheobronchitis, which is characterized by a "barking" cough but can also include laryngitis.

Beyond viruses, laryngitis can be caused in rare cases by bacteria and fungi. Bacterial laryngitis can be a complication of viral laryngitis, occurring about 7 days into the illness. The most common bacteria involved are *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*. Fungal laryngitis is probably rarer but should be considered in patients who are immunosuppressed or who have recently been treated with anti-bacterial drugs.

Noninfectious causes of laryngitis include vocal trauma (e.g., due to yelling, screaming, or loud singing), inhalation injuries, allergies, gastroesophageal reflux disease (laryngopharyngeal reflux), asthma, and pollution. Immunocompetent patients are at risk for infections with herpesvirus, HIV, and coxsackievirus. Smokers are at elevated risk for malignancy and other infections.

Laryngitis is characterized by a raspy, hoarse, or breathy voice, sometimes progressing to a complete loss of voice. Laryngitis can have associated dry cough and anterior throat pain; patients often feel a need to clear their throats. The physical examination in patients who may have laryngitis should focus on the head, neck, and lungs, but the diagnosis of laryngitis is generally based on history. If visualization of the vocal cords is necessary, indirect examination with a mirror or flexible laryngoscopy usually shows erythema and edema of the vocal cords and surrounding structures.

TREATMENT

Laryngitis

Laryngitis is generally self-limited, usually lasting 3–7 days, but may last up to 14 days. Vocal rest is crucial. Airway humidification and hydration should help. Patients likely to have laryngopharyngeal reflux should avoid gastroesophageal reflux-inducing foods and behaviors and should take antireflux medications. In randomized controlled trials, antibiotics were not effective in decreasing objective symptoms of laryngitis.

Red flags for emergency evaluation and monitoring include shortness of breath, stridor, dysphagia, odynophagia, drooling, and posturing that could indicate epiglottitis. Referral to an otolaryngologist should be considered for patients who rely on their voice for work, such as singers and teachers. A history of smoking or weight loss should raise suspicion of malignancy. Symptoms lasting >3 weeks should prompt referral to an otolaryngologist or speech specialist.

■ FURTHER READING

- CENTOR RM, LINDER JA: Web exclusive. Annals on call—*Fusobacterium* pharyngitis debate. Ann Intern Med 171:OC1, 2019.
- CHUA KP et al: Appropriateness of outpatient antibiotic prescribing among privately insured US patients: ICD-10-CM based cross sectional study. BMJ 364:k5092, 2019.
- LIEBERTHAL AS et al: Clinical practice guideline: The diagnosis and management of acute otitis media. Pediatrics 131:e964, 2013.
- ROWE TA, LINDER JA: Novel approaches to decrease inappropriate ambulatory antibiotic use. Expert Rev Anti Infect Ther 17:511, 2019.
- SANCHEZ GV et al: Core elements of outpatient antibiotic stewardship. MMWR Recomm Rep 65:1, 2016.



As primary care physicians and consultants, internists are often asked to evaluate patients with disease of the oral soft tissues, teeth, and pharynx. Knowledge of the oral milieu and its unique structures is necessary to guide preventive services and recognize oral manifestations of local or systemic disease (Chap. A3). Furthermore, internists frequently collaborate with dentists in the care of patients who have a variety of medical conditions that affect oral health or who undergo dental procedures that increase their risk of medical complications.

DISEASES OF THE TEETH AND PERIODONTAL STRUCTURES

Tooth formation begins during the sixth week of embryonic life and continues through 17 years of age. Teeth start to develop in utero and continue to develop until after the tooth erupts. Normally, all 20 deciduous teeth have erupted by age 3 and have been shed by age 13. Permanent teeth, eventually totaling 32, begin to erupt by age 6 and have completely erupted by age 14, though third molars ("wisdom teeth") may erupt later.

The erupted tooth consists of the visible *crown* covered with enamel and the root submerged below the gum line and covered with bonelike *cementum*. *Dentin*, a material that is denser than bone and exquisitely sensitive to pain, forms the majority of the tooth substance, surrounding a core of myxomatous *pulp* containing the vascular and nerve supply. The tooth is held firmly in the alveolar socket by the *periodontium*, supporting structures that consist of the gingivae, alveolar bone, cementum, and periodontal ligament. The periodontal ligament tenaciously binds the tooth's cementum to the alveolar bone. Above this ligament is a collar of attached gingiva just below the crown. A few millimeters of unattached or free gingiva (1–3 mm) overlap the base of the crown, forming a shallow sulcus along the gum-tooth margin.

Dental Caries, Pulpal and Periapical Disease, and Complications Dental caries usually begin asymptotically as a destructive infectious process of the enamel. Bacteria—principally *Streptococcus mutans*—colonize the organic buffering biofilm (*plaque*) on the tooth surface. If not removed by brushing or by the natural cleansing and antibacterial action of saliva, bacterial acids can demineralize the enamel. Fissures and pits on the occlusal surfaces are the most frequent sites of early decay. Surfaces between the teeth, adjacent to tooth restorations and exposed roots, are also vulnerable, particularly as individuals age. Over time, dental caries extend to the underlying dentin, leading to cavitation of the enamel. Without management, the caries will penetrate to the tooth pulp, producing *acute pulpitis*. At this stage, when the pulp infection is limited, the tooth may become sensitive to percussion and to hot or cold, and pain resolves immediately when the irritating stimulus is removed. Should the infection spread throughout the pulp, *irreversible pulpitis* occurs, leading to *pulp necrosis*. At this later stage, pain can be severe and has a sharp or throbbing visceral quality that may be worse when the patient lies down. Once pulp necrosis is complete, pain may be constant or intermittent, but cold sensitivity is lost.

Treatment of caries involves removal of the softened and infected hard tissue and restoration of the tooth structure with silver amalgam, glass ionomer, composite resin, or gold. Once irreversible pulpitis occurs, root canal therapy becomes necessary; removal of the contents of the pulp chamber and root canal is followed by thorough cleaning and filling with an inert material. Alternatively, the tooth may be extracted.

Pulpal infection leads to *periapical abscess* formation, which can produce pain on chewing. If the infection is mild and chronic, a *periapical granuloma* or eventually a *periapical cyst* forms, either of which

produces radiolucency at the root apex. When unchecked, a periapical abscess can erode into the alveolar bone, producing osteomyelitis; penetrate and drain through the gingivae, producing a parulis (gum-boil); or track along deep fascial planes, producing virulent cellulitis (Ludwig's angina) involving the submandibular space and floor of the mouth (Chap. 177). Elderly patients, patients with diabetes mellitus, and patients taking glucocorticoids may experience little or no pain or fever as these complications develop.

Periodontal Disease Periodontal disease and dental caries are the primary causes of tooth loss. Like dental caries, chronic infection of the gingiva and anchoring structures of the tooth begins with formation of bacterial plaque. The process begins at the gum line. Plaque and *calculus* (calcified plaque) are preventable by appropriate daily oral hygiene, including periodic professional cleaning. Left undisturbed, chronic inflammation can ensue and produce hyperemia of the free and attached gingivae (*gingivitis*), which then typically bleed with brushing. If this issue is ignored, severe *periodontitis* can develop, leading to deepening of the physiologic sulcus and destruction of the periodontal ligament. Gingival pockets develop around the teeth. As the periodontium (including the supporting bone) is destroyed, the teeth loosen. A role for chronic inflammation due to chronic periodontal disease in promoting coronary heart disease and stroke has been proposed. Epidemiologic studies have demonstrated a moderate but significant association between chronic periodontal inflammation and atherosclerosis, though a causal role remains unproven.

Acute and aggressive forms of periodontal disease are less common than the chronic forms described above. However, if the host is stressed or exposed to a new pathogen, rapidly progressive and destructive disease of the periodontal tissue can occur. A virulent example is *acute necrotizing ulcerative gingivitis*. The presentation includes sudden gingival inflammation, ulceration, bleeding, interdental gingival necrosis, and fetid halitosis. *Localized juvenile periodontitis*, which is seen in adolescents, is particularly destructive and appears to be associated with impaired neutrophil chemotaxis. *AIDS-related periodontitis* resembles acute necrotizing ulcerative gingivitis in some patients and a more destructive form of adult chronic periodontitis in others. It may also produce a gangrene-like destructive process of the oral soft tissues and bone that resembles *noma*, an infectious condition seen in severely malnourished children in developing nations.

Prevention of Tooth Decay and Periodontal Infection Despite the reduced prevalences of dental caries and periodontal disease in the United States (due in large part to water fluoridation and improved dental care, respectively), both diseases constitute a major public health problem worldwide, particularly in certain groups. The internist should promote preventive dental care and hygiene as part of health maintenance. Populations at high risk for dental caries and periodontal disease include those with hyposalivation and/or xerostomia, diabetics, alcoholics, tobacco users, persons with Down syndrome, and those with gingival hyperplasia. Furthermore, patients lacking access to dental care (e.g., as a result of low socioeconomic status) and patients with a reduced ability to provide self-care (e.g., individuals with disabilities, nursing home residents, and persons with dementia or upper-extremity disability) suffer at a disproportionate rate. It is important to provide counseling regarding regular dental hygiene and professional cleaning, use of fluoride-containing toothpaste, professional fluoride treatments, and (for patients with limited dexterity) use of electric toothbrushes and also to instruct persons caring for those who are not capable of self-care. Cost, fear of dental care, and differences in language and culture create barriers that prevent some people from seeking preventive dental services.

Developmental and Systemic Disease Affecting the Teeth and Periodontium In addition to posing cosmetic issues, *malocclusion*, the most common developmental oral problem, can interfere with mastication unless corrected through orthodontic and surgical techniques. Impacted third molars are common and can become infected or erupt into an insufficient space. Acquired prognathism due to *acromegaly* may also lead to malocclusion, as may deformity of the maxilla and

mandible due to *Paget's disease* of the bone. Delayed tooth eruption, a receding chin, and a protruding tongue are occasional features of *cretinism* and *hypopituitarism*. Congenital syphilis produces tapering, notched (*Hutchinson's*) incisors and finely nodular (*mulberry*) molar crowns. *Enamel hypoplasia* results in crown defects ranging from pits to deep fissures of primary or permanent teeth. Intrauterine infection (syphilis, rubella), vitamin deficiency (A, C, or D), disorders of calcium metabolism (malabsorption, vitamin D-resistant rickets, hypoparathyroidism), prematurity, high fever, and rare inherited defects (*amelogenesis imperfecta*) are all causes. Tetracycline, given in sufficiently high doses during the first 8 years of life, may produce enamel hypoplasia and discoloration. Doxycycline does not cause permanent tooth staining in children despite warnings included for all tetracycline-class antibiotics. Exposure to endogenous pigments can discolor developing teeth; etiologies include *erythroblastosis fetalis* (green or bluish-black), congenital liver disease (green or yellow-brown), and porphyria (red or brown that fluoresces with ultraviolet light). *Mottled enamel* occurs if excessive fluoride is ingested during development. Worn enamel is seen with age, bruxism, or excessive acid exposure (e.g., chronic gastric reflux or bulimia). Celiac disease is associated with nonspecific enamel defects in children but not in adults.

Total or partial tooth loss resulting from periodontitis is seen with cyclic neutropenia, Papillon-Lefèvre syndrome, Chédiak-Higashi syndrome, and leukemia. Rapid focal tooth loosening is most often due to infection, but rarer causes include Langerhans cell histiocytosis, Ewing's sarcoma, osteosarcoma, and Burkitt's lymphoma. Early loss of primary teeth is a feature of *hypophosphatasia*, a rare congenital error of metabolism.

Pregnancy may produce gingivitis and localized *pyogenic granulomas*. Severe periodontal disease occurs in uncontrolled diabetes mellitus. *Drug-induced gingival overgrowth* may be caused by anti-convulsants, calcium channel blockers, and immunosuppressants, although excellent daily oral care can prevent or reduce its occurrence. *Idiopathic familial gingival fibromatosis* and several syndrome-related disorders cause similar conditions. Discontinuation of the medication may reverse the drug-induced form, although surgery may be needed to control both of the latter entities. *Linear gingival erythema* is variably seen in patients with advanced HIV infection and probably represents immune deficiency and decreased neutrophil activity. Diffuse or focal gingival swelling may be a feature of early or late acute myelomonocytic leukemia as well as of other lymphoproliferative disorders. A rare but pathognomonic sign of granulomatosis with polyangiitis is a red-purple, granular gingivitis (*strawberry gums*).

DISEASES OF THE ORAL MUCOSA

Infections Most oral mucosal diseases involve microorganisms (Table 36-1).

Pigmented Lesions See Table 36-2.

Dermatologic Diseases See Tables 36-1, 36-2, and 36-3 and Chaps. 56–61.

Diseases of the Tongue See Table 36-4.

HIV Disease and AIDS See Tables 36-1, 36-2, 36-3, and 36-5; Chap. 202.

Ulcers Ulceration is the most common oral mucosal lesion. Although there are many causes, the host and the pattern of lesions, including the presence of organ system features, narrow the differential diagnosis (Table 36-1). Most acute ulcers are painful and self-limited. Recurrent aphthous ulcers and herpes simplex account for the majority. Persistent and deep aphthous ulcers can be idiopathic or can accompany HIV/AIDS. Aphthous lesions are often the presenting symptom in *Behcet's syndrome* (Chap. 364). Similar-appearing, though less painful, lesions may occur in reactive arthritis, and aphthous ulcers are occasionally present during phases of *discoid* or *systemic lupus erythematosus* (Chap. 360). Aphthous-like ulcers are seen in *Crohn's disease* (Chap. 326), but, unlike the common aphthous variety, they

may exhibit granulomatous inflammation on histologic examination. Recurrent aphthae are more prevalent in patients with *celiac disease* and have been reported to remit with elimination of gluten.

Of major concern are chronic, relatively painless ulcers and mixed red/white patches (erythroplakia and leukoplakia) of >2 weeks' duration. Squamous cell carcinoma and premalignant dysplasia should be considered early and a diagnostic biopsy performed. This awareness and this procedure are critically important because early-stage malignancy is vastly more treatable than late-stage disease. High-risk sites include the lower lip, floor of the mouth, ventral and lateral tongue, and soft palate-tonsillar pillar complex. Significant risk factors for oral cancer in Western countries include sun exposure (lower lip), tobacco and alcohol use, and human papillomavirus infection. In India and some other Asian countries, smokeless tobacco mixed with betel nut, slaked lime, and spices is a common cause of oral cancer. Rarer causes of chronic oral ulcer, such as tuberculosis, fungal infection, granulomatosis with polyangiitis, and midline granuloma, may look identical to carcinoma. Making the correct diagnosis depends on recognizing other clinical features and performing a biopsy of the lesion. The syphilitic chancre is typically painless and therefore easily missed. Regional lymphadenopathy is invariably present. The syphilitic etiology is confirmed with appropriate bacterial and serologic tests.

Disorders of mucosal fragility often produce painful oral ulcers that fail to heal within 2 weeks. *Mucous membrane pemphigoid* and *pemphigus vulgaris* are the major acquired disorders. While their clinical features are often distinctive, a biopsy or immunohistochemical examination should be performed to diagnose these entities and to distinguish them from *lichen planus* and drug reactions.

Hematologic and Nutritional Disease Internists are more likely to encounter patients with acquired, rather than congenital, bleeding disorders. Bleeding should stop 15 min after minor trauma and within an hour after tooth extraction if local pressure is applied. More prolonged bleeding, if not due to continued injury or rupture of a large vessel, should lead to investigation for a clotting abnormality. In addition to bleeding, petechiae and ecchymoses are prone to occur at the vibrating line between the soft and hard palates in patients with platelet dysfunction or thrombocytopenia.

All forms of leukemia, but particularly *acute myelomonocytic leukemia*, can produce gingival bleeding, ulcers, and gingival enlargement. Oral ulcers are a feature of agranulocytosis, and ulcers and mucositis are often severe complications of chemotherapy and radiation therapy for hematologic and other malignancies. *Plummer-Vinson syndrome* (iron deficiency, angular stomatitis, glossitis, and dysphagia) raises the risk of oral squamous cell cancer and esophageal cancer at the postcricoidal tissue web. Atrophic papillae and a red, burning tongue may occur with pernicious anemia. Deficiencies in B-group vitamins produce many of these same symptoms as well as oral ulceration and cheilosis. Consequences of *scurvy* include swollen, bleeding gums; ulcers; and loosening of the teeth.

NONDENTAL CAUSES OF ORAL PAIN

Most, but not all, oral pain emanates from inflamed or injured tooth pulp or periodontal tissues. Nonodontogenic causes are often overlooked. In most instances, toothache is predictable and proportional to the stimulus applied, and an identifiable condition (e.g., caries, abscess) is found. Local anesthesia eliminates pain originating from dental or periodontal structures, but not referred pains. The most common nondental source of pain is myofascial pain referred from muscles of mastication, which become tender and ache with increased use. Many sufferers exhibit *bruxism* (grinding of the teeth) secondary to stress and anxiety. *Temporomandibular joint disorder* is closely related. It affects both sexes, with a higher prevalence among women. Features include pain, limited mandibular movement, and temporomandibular joint sounds. The etiologies are complex; malocclusion does not play the primary role once attributed to it. *Osteoarthritis* is a common cause of masticatory pain. Anti-inflammatory medication, jaw rest, soft foods, and heat provide relief. The temporomandibular joint is involved in 50% of patients with *rheumatoid arthritis*, and its involvement is

TABLE 36-1 Vesicular, Bullous, or Ulcerative Lesions of the Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Viral Diseases			
Primary acute herpetic gingivostomatitis (HSV type 1; rarely type 2)	Lip and oral mucosa (buccal, gingival, lingual mucosa)	Labial vesicles that rupture and crust, and intraoral vesicles that quickly ulcerate; extremely painful; acute gingivitis, fever, malaise, foul odor, and cervical lymphadenopathy; occurs primarily in infants, children, and young adults	Heals spontaneously in 10–14 days; unless secondarily infected, lesions lasting >3 weeks are not due to primary HSV infection
Recurrent herpes labialis	Mucocutaneous junction of lip, perioral skin	Eruption of groups of vesicles that may coalesce, then rupture and crust; painful to pressure or spicy foods	Lasts ~1 week, but condition may be prolonged if secondarily infected; if severe, topical or oral antiviral treatment may reduce healing time
Recurrent intraoral herpes simplex	Palate and gingiva	Small vesicles on keratinized epithelium that rupture and coalesce; painful	Heals spontaneously in ~1 week; if severe, topical or oral antiviral treatment may reduce healing time
Chickenpox (VZV)	Gingiva and oral mucosa	Skin lesions may be accompanied by small vesicles on oral mucosa that rupture to form shallow ulcers; may coalesce to form large bullous lesions that ulcerate; mucosa may have generalized erythema	Lesions heal spontaneously within 2 weeks
Herpes zoster (VZV reactivation)	Cheek, tongue, gingiva, or palate	Unilateral vesicular eruptions and ulceration in linear pattern following sensory distribution of trigeminal nerve or one of its branches	Gradual healing without scarring unless secondarily infected; postherpetic neuralgia is common; oral acyclovir, famciclovir, or valacyclovir reduces healing time and postherpetic neuralgia
Infectious mononucleosis (Epstein-Barr virus)	Oral mucosa	Fatigue, sore throat, malaise, fever, and cervical lymphadenopathy; numerous small ulcers usually appear several days before lymphadenopathy; gingival bleeding and multiple petechiae at junction of hard and soft palates	Oral lesions disappear during convalescence; no treatment is given, though glucocorticoids are indicated if tonsillar swelling compromises the airway
Herpangina (coxsackievirus A; also possibly coxsackievirus B and echovirus)	Oral mucosa, pharynx, tongue	Sudden onset of fever, sore throat, and oropharyngeal vesicles, usually in children <4 years old, during summer months; diffuse pharyngeal congestion and vesicles (1–2 mm), grayish-white surrounded by red areola; vesicles enlarge and ulcerate	Incubation period of 2–9 days; fever for 1–4 days; recovery uneventful
Hand-foot-and-mouth disease (most commonly coxsackievirus A16)	Oral mucosa, pharynx, palms, and soles	Fever, malaise, headache with oropharyngeal vesicles that become painful, shallow ulcers; highly infectious; usually affects children under age 10	Incubation period 2–18 days; lesions heal spontaneously in 2–4 weeks
Primary HIV infection	Gingiva, palate, and pharynx	Acute gingivitis and oropharyngeal ulceration, associated with febrile illness resembling mononucleosis and including lymphadenopathy	Followed by HIV seroconversion, asymptomatic HIV infection, and usually ultimately by HIV disease
Bacterial or Fungal Diseases			
Acute necrotizing ulcerative gingivitis ("trench mouth")	Gingiva	Painful, bleeding gingiva characterized by necrosis and ulceration of gingival papillae and margins plus lymphadenopathy and foul breath	Debridement and diluted (1:3) peroxide lavage provide relief within 24 h; antibiotics in acutely ill patients; relapse may occur
Prenatal (congenital) syphilis	Palate, jaws, tongue, and teeth	Gummatus involvement of palate, jaws, and facial bones; Hutchinson's incisors, mulberry molars, glossitis, mucous patches, and fissures at corner of mouth	Tooth deformities in permanent dentition irreversible
Primary syphilis (chancre)	Lesion appearing where organism enters body; may occur on lips, tongue, or tonsillar area	Small papule developing rapidly into a large, painless ulcer with indurated border; unilateral lymphadenopathy; chancre and lymph nodes containing spirochetes; serologic tests positive by third to fourth weeks	Healing of chancre in 1–2 months, followed by secondary syphilis in 6–8 weeks
Secondary syphilis	Oral mucosa frequently involved with mucous patches, which occur primarily on palate and also at commissures of mouth	Maculopapular lesions of oral mucosa, 5–10 mm in diameter with central ulceration covered by grayish membrane; eruptions occurring on various mucosal surfaces and skin, accompanied by fever, malaise, and sore throat	Lesions may persist from several weeks to a year
Tertiary syphilis	Palate and tongue	Gummatus infiltration of palate or tongue followed by ulceration and fibrosis; atrophy of tongue papillae produces characteristic bald tongue and glossitis	Gumma may destroy palate, causing complete perforation
Gonorrhea	Lesions may occur in mouth at site of inoculation or secondarily by hematogenous spread from a primary focus	Most pharyngeal infection is asymptomatic; may produce burning or itching sensation; oropharynx and tonsils may be ulcerated and erythematous; saliva viscous and fetid	More difficult to eradicate than urogenital infection, though pharyngitis usually resolves with appropriate antimicrobial treatment
Tuberculosis	Tongue, tonsillar area, soft palate	Painless, solitary, 1- to 5-cm, irregular ulcer covered with persistent exudate; ulcer has firm undermined border	Autoinoculation from pulmonary infection is usual; lesions resolve with appropriate antimicrobial therapy
Cervicofacial actinomycosis	Swellings in region of face, neck, and floor of mouth	Infection may be associated with extraction, jaw fracture, or eruption of molar tooth; in acute form, resembles acute pyogenic abscess, but contains yellow "sulfur granules" (gram-positive mycelia and their hyphae)	Typically, swelling is hard and grows painlessly; multiple abscesses with draining tracts develop; penicillin first choice; surgery usually necessary

(Continued)

TABLE 36-1 Vesicular, Bullous, or Ulcerative Lesions of the Oral Mucosa (Continued)

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Bacterial or Fungal Diseases (Continued)			
Histoplasmosis	Any area of the mouth, particularly tongue, gingiva, or palate	Nodular, verrucous, or granulomatous lesions; ulcers are indurated and painful; usual source hematogenous or pulmonary, but may be primary	Systemic antifungal therapy necessary
Candidiasis^a			
Dermatologic Diseases			
Mucous membrane pemphigoid	Typically produces marked gingival erythema and ulceration; other areas of oral cavity, esophagus, and vagina may be affected	Painful, grayish-white collapsed vesicles or bullae of full-thickness epithelium with peripheral erythematous zone; gingival lesions desquamate, leaving ulcerated area	Protracted course with remissions and exacerbations; involvement of different sites develops slowly; glucocorticoids may temporarily reduce symptoms but do not control disease
EM minor and EM major (Stevens-Johnson syndrome)	Primarily oral mucosa and skin of hands and feet	Intraoral ruptured bullae surrounded by inflammatory area; lips may show hemorrhagic crusts; "iris" or "target" lesion on skin is pathognomonic; patient may have severe signs of toxicity	Onset very rapid; usually idiopathic, but may be associated with trigger such as drug reaction; condition may last 3–6 weeks; mortality rate for untreated EM major is 5–15%
Pemphigus vulgaris	Oral mucosa and skin; sites of mechanical trauma (soft/hard palate, frenulum, lips, buccal mucosa)	Usually (>70%) presents with oral lesions; fragile, ruptured bullae and ulcerated oral areas; mostly in older adults	With repeated occurrence of bullae, toxicity may lead to cachexia, infection, and death within 2 years; often controllable with oral glucocorticoids
Lichen planus	Oral mucosa and skin	White striae in mouth; purplish nodules on skin at sites of friction; occasionally causes oral mucosal ulcers and erosive gingivitis	White striae alone usually asymptomatic; erosive lesions often difficult to treat, but may respond to glucocorticoids
Other Conditions			
Recurrent aphthous ulcers	Usually on nonkeratinized oral mucosa (buccal and labial mucosa, floor of mouth, soft palate, lateral and ventral tongue)	Single or clustered painful ulcers with surrounding erythematous border; lesions may be 1–2 mm in diameter in crops (herpetiform), 1–5 mm (minor), or 5–15 mm (major)	Lesions heal in 1–2 weeks but may recur monthly or several times a year; protective barrier with benzocaine and topical glucocorticoids relieve symptoms; systemic glucocorticoids may be needed in severe cases
Behçet's syndrome	Oral mucosa, eyes, genitalia, gut, and CNS	Multiple aphthous ulcers in mouth; inflammatory ocular changes, ulcerative lesions on genitalia; inflammatory bowel disease and CNS disease	Oral lesions often first manifestation; persist several weeks and heal without scarring
Traumatic ulcers	Anywhere on oral mucosa; dentures frequently responsible for ulcers in vestibule	Localized, discrete ulcerated lesions with red border; produced by accidental biting of mucosa, penetration by foreign object, or chronic irritation by dentures	Lesions usually heal in 7–10 days when irritant is removed, unless secondarily infected
Squamous cell carcinoma	Any area of mouth, most commonly on lower lip, lateral borders of tongue, and floor of mouth	Red, white, or red and white ulcer with elevated or indurated border; failure to heal; pain not prominent in early lesions	Invades and destroys underlying tissues; frequently metastasizes to regional lymph nodes
Acute myeloid leukemia (usually monocytic)	Gingiva	Gingival swelling and superficial ulceration followed by hyperplasia of gingiva with extensive necrosis and hemorrhage; deep ulcers may occur elsewhere on mucosa, complicated by secondary infection	Usually responds to systemic treatment of leukemia; occasionally requires local irradiation
Lymphoma	Gingiva, tongue, palate, and tonsillar area	Elevated, ulcerated area that may proliferate rapidly, giving appearance of traumatic inflammation	Fatal if untreated; may indicate underlying HIV infection
Chemical or thermal burns	Any area in mouth	White slough due to contact with corrosive agents (e.g., aspirin, hot cheese) applied locally; removal of slough leaves raw, painful surface	Lesion heals in several weeks if not secondarily infected

^aSee Table 36-3.

Abbreviations: CNS, central nervous system; EM, erythema multiforme; HSV, herpes simplex virus; VZV, varicella-zoster virus.

usually a late feature of severe disease. Bilateral preauricular pain, particularly in the morning, limits range of motion.

Migrainous neuralgia may be localized to the mouth. Episodes of pain and remission without an identifiable cause and a lack of relief with local anesthesia are important clues. *Trigeminal neuralgia (tic dououreux)* can involve the entire branch or part of the mandibular or maxillary branch of the fifth cranial nerve and can produce pain in one or a few teeth. Pain may occur spontaneously or may be triggered by touching the lip or gingiva, brushing the teeth, or chewing. *Glossopharyngeal neuralgia* produces similar acute neuropathic symptoms

in the distribution of the ninth cranial nerve. Swallowing, sneezing, coughing, or pressure on the tragus of the ear triggers pain that is felt in the base of the tongue, pharynx, and soft palate and may be referred to the temporomandibular joint. *Neuritis* involving the maxillary and mandibular divisions of the trigeminal nerve (e.g., maxillary sinusitis, neuroma, and leukemic infiltrate) is distinguished from ordinary toothache by the neuropathic quality of the pain. Occasionally, *phantom pain* follows tooth extraction. Pain and hyperalgesia behind the ear and on the side of the face in the day or so before facial weakness develops often constitute the earliest symptom of *Bell's palsy*. Likewise,

TABLE 36-2 Pigmented Lesions of the Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Oral melanotic macule	Any area of mouth	Discrete or diffuse, localized, brown to black macule	Remains indefinitely; no growth
Diffuse melanin pigmentation	Any area of mouth	Diffuse pale to dark-brown pigmentation; may be physiologic ("racial") or due to smoking	Remains indefinitely
Nevi	Any area of mouth	Discrete, localized, brown to black pigmentation	Remains indefinitely
Malignant melanoma	Any area of mouth	Can be flat and diffuse, painless, brown to black; or can be raised and nodular	Expands and invades early; metastasis leads to death
Addison's disease	Any area of mouth, but mostly buccal mucosa	Blotches or spots of bluish-black to dark-brown pigmentation occurring early in disease, accompanied by diffuse pigmentation of skin; other symptoms of adrenal insufficiency	Condition controlled by adrenal steroid replacement
Peutz-Jeghers syndrome	Any area of mouth	Dark-brown spots on lips, buccal mucosa, with characteristic distribution of pigment around lips, nose, and eyes and on hands; concomitant intestinal polyposis	Oral pigmented lesions remain indefinitely; gastrointestinal polyps may become malignant
Drug ingestion (neuroleptics, oral contraceptives, minocycline, zidovudine, quinine derivatives)	Any area of mouth	Brown, black, or gray areas of pigmentation	Gradually disappears following cessation of drug intake
Amalgam tattoo	Gingiva and alveolar mucosa	Small blue-black pigmented areas associated with embedded amalgam particles in soft tissues; may show up on radiographs as radiopaque particles in some cases	Remains indefinitely
Heavy metal pigmentation (bismuth, mercury, lead)	Gingival margin	Thin blue-black pigmented line along gingival margin; rarely seen except in children exposed to lead-based paint	Indicative of systemic absorption; no significance for oral health
Black hairy tongue	Dorsum of tongue	Elongation of filiform papillae of tongue, which become stained by coffee, tea, tobacco, or pigmented bacteria	Improves within 1–2 weeks with gentle brushing of tongue or (if due to bacterial overgrowth) discontinuation of antibiotic
Fordyce spots	Buccal and labial mucosa	Numerous small yellowish spots just beneath mucosal surface; no symptoms; due to hyperplasia of sebaceous glands	Benign; remains without apparent change
Kaposi's sarcoma	Palate most common, but may occur at any other site	Red or blue plaques of variable size and shape; often enlarge, become nodular, and may ulcerate	Usually indicative of HIV infection or non-Hodgkin's lymphoma; rarely fatal, but may require treatment for comfort or cosmesis
Mucous retention cysts	Buccal and labial mucosa	Bluish, clear fluid-filled cyst due to extravasated mucus from injured minor salivary gland	Benign; painless unless traumatized; may be removed surgically

TABLE 36-3 White Lesions of Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Lichen planus	Buccal mucosa, tongue, gingiva, and lips; skin	Striae, white plaques, red areas, ulcers in mouth; purplish papules on skin; may be asymptomatic, sore, or painful; lichenoid drug reactions may look similar	Protracted; responds to topical glucocorticoids
White sponge nevus	Oral mucosa, vagina, anal mucosa	Painless white thickening of epithelium; adolescence/early adulthood onset; familial	Benign and permanent
Smoker's leukoplakia and smokeless tobacco lesions	Any area of oral mucosa, sometimes related to location of habit	White patch that may become firm, rough, or red-fissured and ulcerated; may become sore and painful but is usually painless	May or may not resolve with cessation of habit; 2% of patients develop squamous cell carcinoma; early biopsy essential
Erythroplakia with or without white patches	Floor of mouth commonly affected in men; tongue and buccal mucosa in women	Velvety, reddish plaque; occasionally mixed with white patches or smooth red areas	High risk of squamous cell cancer; early biopsy essential
Candidiasis	Any area in mouth	<i>Pseudomembranous type</i> ("thrush"): creamy white curdlike patches that reveal a raw, bleeding surface when scraped; found in sick infants, debilitated elderly patients receiving high-dose glucocorticoids or broad-spectrum antibiotics, and patients with AIDS	Responds favorably to antifungal therapy and correction of predisposing causes where possible
		<i>Erythematous type</i> : flat, red, sometimes sore areas in same groups of patients	Course same as for pseudomembranous type
		<i>Candidal leukoplakia</i> : nonremovable white thickening of epithelium due to <i>Candida</i>	Responds to prolonged antifungal therapy
		<i>Angular cheilitis</i> : sore fissures at corner of mouth	Responds to topical antifungal therapy
Hairy leukoplakia	Usually on lateral tongue, rarely elsewhere on oral mucosa	White areas ranging from small and flat to extensive accentuation of vertical folds; found in HIV carriers (all risk groups for AIDS)	Due to Epstein-Barr virus; responds to high-dose acyclovir but recurs; rarely causes discomfort unless secondarily infected with <i>Candida</i>
Warts (human papillomavirus)	Anywhere on skin and oral mucosa	Single or multiple papillary lesions with thick, white, keratinized surfaces containing many pointed projections; cauliflower lesions covered with normal-colored mucosa or multiple pink or pale bumps (focal epithelial hyperplasia)	Lesions grow rapidly and spread; squamous cell carcinoma must be ruled out with biopsy; excision or laser therapy; may regress in HIV-infected patients receiving antiretroviral therapy

TABLE 36-4 Alterations of the Tongue

TYPE OF CHANGE	CLINICAL FEATURES
Size or Morphology	
Macroglossia	Enlarged tongue that may be part of a syndrome found in developmental conditions such as Down syndrome, Simpson-Golabi-Behmel syndrome, or Beckwith-Wiedemann syndrome; may be due to tumor (hemangioma or lymphangioma), metabolic disease (e.g., primary amyloidosis), or endocrine disturbance (e.g., acromegaly or cretinism); may occur when all teeth are removed
Fissured ("scrotal") tongue	Dorsal surface and sides of tongue covered by painless shallow or deep fissures that may collect debris and become irritated
Median rhomboid glossitis	Congenital abnormality with ovoid, denuded area in median posterior portion of tongue; may be associated with candidiasis and may respond to antifungal treatment
Color	
"Geographic" tongue (benign migratory glossitis)	Asymptomatic inflammatory condition of tongue, with rapid loss and regrowth of filiform papillae leading to appearance of denuded red patches "wandering" across surface
Hairy tongue	Elongation of filiform papillae of medial dorsal surface area due to failure of keratin layer of papillae to desquamate normally; brownish-black coloration may be due to staining by tobacco, food, or chromogenic organisms
"Strawberry" and "raspberry" tongue	Appearance of tongue during scarlet fever due to hypertrophy of fungiform papillae as well as changes in filiform papillae
"Bald" tongue	Atrophy may be associated with xerostomia, pernicious anemia, iron-deficiency anemia, pellagra, or syphilis; may be accompanied by painful burning sensation; may be an expression of erythematous candidiasis and respond to antifungal treatment

TABLE 36-5 Oral Lesions Associated with HIV Infection

LESION MORPHOLOGY	ETIOLOGIES
Papules, nodules, plaques	Candidiasis (hyperplastic and pseudomembranous) ^a Condyloma acuminatum (human papillomavirus infection) Squamous cell carcinoma (preinvasive and invasive) Non-Hodgkin's lymphoma ^a Hairy leukoplakia ^a
Ulcers	Recurrent aphthous ulcers ^a Angular cheilitis Squamous cell carcinoma Acute necrotizing ulcerative gingivitis ^a Necrotizing ulcerative periodontitis ^a Necrotizing ulcerative stomatitis Non-Hodgkin's lymphoma ^a Viral infection (herpes simplex, herpes zoster, cytomegalovirus infection) Infection caused by <i>Mycobacterium tuberculosis</i> or <i>Mycobacterium avium-intracellulare</i> Fungal infection (histoplasmosis, cryptococcosis, candidiasis, geotrichosis, aspergillosis) Bacterial infection (<i>Escherichia coli</i> , <i>Enterobacter cloacae</i> , <i>Klebsiella pneumoniae</i> , <i>Pseudomonas aeruginosa</i>) Drug reactions (single or multiple ulcers)
Pigmented lesions	Kaposi's sarcoma ^a Bacillary angiomatosis (skin and visceral lesions more common than oral) Zidovudine pigmentation (skin, nails, and occasionally oral mucosa) Addison's disease
Miscellaneous	Linear gingival erythema ^a

^aStrongly associated with HIV infection.

similar symptoms may precede visible lesions of herpes zoster infecting the seventh nerve (*Ramsey-Hunt syndrome*) or trigeminal nerve. *Postherpetic neuralgia* may follow either condition. *Coronary ischemia* may produce pain exclusively in the face and jaw; as in typical angina pectoris, this pain is usually reproducible with increased myocardial demand. Aching in several upper molar or premolar teeth that is unrelieved by anesthetizing the teeth may point to *maxillary sinusitis*.

Giant cell arteritis is notorious for producing headache, but it may also produce facial pain or sore throat without headache. Jaw and tongue claudication with chewing or talking is relatively common. Tongue infarction is rare. Patients with subacute thyroiditis often experience pain referred to the face or jaw before the tenderness of the thyroid gland and transient hyperthyroidism are appreciated.

"Burning mouth syndrome" (*glossodynia*) occurs in the absence of an identifiable cause (e.g., vitamin B₁₂ deficiency, iron deficiency, diabetes mellitus, low-grade *Candida* infection, food sensitivity, or subtle xerostomia) and predominantly affects postmenopausal women. The etiology may be neuropathic. Clonazepam, α-lipoic acid, and cognitive-behavioral therapy have benefited some patients. Some cases associated with an angiotensin-converting enzyme inhibitor have remitted when treatment with the drug was discontinued.

DISEASES OF THE SALIVARY GLANDS

Saliva is essential to oral health. Its absence leads to dental caries, periodontal disease, and difficulties in wearing dental prostheses, masticating, and speaking. Its major components, water and mucin, serve as a cleansing solvent and lubricating fluid. In addition, saliva contains antimicrobial factors (e.g., lysozyme, lactoperoxidase, secretory IgA), epidermal growth factor, minerals, and buffering systems. The major salivary glands secrete intermittently in response to autonomic stimulation, which is high during a meal but low otherwise. Hundreds

of minor glands in the lips and cheeks secrete mucus continuously throughout the day and night. Consequently, oral function becomes impaired when salivary function is reduced. The sensation of a dry mouth (*xerostomia*) is perceived when salivary flow is reduced by 50%. The most common etiology is medication, especially drugs with anticholinergic properties but also alpha and beta blockers, calcium channel blockers, and diuretics. Other causes include Sjögren's syndrome, chronic parotitis, salivary duct obstruction, diabetes mellitus, HIV/AIDS, and radiation therapy that includes the salivary glands in the field (e.g., for Hodgkin's lymphoma and for head and neck cancer). Management involves the elimination or limitation of drying medications, preventive dental care, and supplementation with oral liquid or salivary substitutes. Sugarless mints or chewing gum may stimulate salivary secretion if dysfunction is mild. When sufficient exocrine tissue remains, pilocarpine or cevimeline has been shown to increase secretions. Commercial saliva substitutes or gels relieve dryness. Fluoride supplementation is critical to prevent caries.

Sialolithiasis presents most often as painful swelling but in some instances as only swelling or only pain. Conservative therapy consists of local heat, massage, and hydration. Promotion of salivary secretion with mints or lemon drops may flush out small stones. Antibiotic treatment is necessary when bacterial infection is suspected. In adults, *acute bacterial parotitis* is typically unilateral and most commonly affects postoperative, dehydrated, and debilitated patients. *Staphylococcus aureus* (including methicillin-resistant strains) and anaerobic bacteria are the most common pathogens. Chronic bacterial *sialadenitis* results from lowered salivary secretion and recurrent bacterial infection. When suspected bacterial infection is not responsive to therapy, the differential diagnosis should be expanded to include benign and malignant neoplasms, lymphoproliferative disorders, Sjögren's syndrome, sarcoidosis, tuberculosis, lymphadenitis, actinomycosis, and

granulomatosis with polyangiitis. Bilateral nontender parotid enlargement occurs with diabetes mellitus, cirrhosis, bulimia, HIV/AIDS, and drugs (e.g., iodide, propylthiouracil).

Pleomorphic adenoma composes two-thirds of all salivary neoplasms. The parotid is the principal salivary gland affected, and the tumor presents as a firm, slow-growing mass. Although this tumor is benign, its recurrence is common if resection is incomplete. Malignant tumors such as mucoepidermoid carcinoma, adenoid cystic carcinoma, and adenocarcinoma tend to grow relatively fast, depending upon grade. They may ulcerate and invade nerves, producing numbness and facial paralysis. Surgical resection is the primary treatment. Radiation therapy (particularly neutron-beam therapy) is used when surgery is not feasible and after resection for certain histologic types with a high risk of recurrence. Malignant salivary gland tumors have a 5-year survival rate of 94% when the stage is local and 35% when distant.

Dental Care for Medically Complex Patients Routine dental care (e.g., uncomplicated extraction, scaling and cleaning, tooth restoration, and root canal) is remarkably safe. The most common concerns regarding care of dental patients with medical disease are excessive bleeding for patients taking anticoagulants, infection of the heart valves and prosthetic devices from hematogenous seeding by the oral flora, and cardiovascular complications resulting from vasopressors used with local anesthetics during dental treatment. Experience confirms that the risk of any of these complications is very low.

Patients undergoing tooth extraction or alveolar and gingival surgery rarely experience uncontrolled bleeding when warfarin anti-coagulation is maintained within the therapeutic range currently recommended for prevention of venous thrombosis, atrial fibrillation, or mechanical heart valve. Embolic complications and death, however, have been reported during subtherapeutic anticoagulation. Therapeutic anticoagulation should be confirmed before and continued through the procedure. Likewise, low-dose aspirin (e.g., 81–325 mg) can safely be continued. For patients taking aspirin and another antiplatelet medication (e.g., clopidogrel), the decision to continue the second antiplatelet medication should be based on individual consideration of the risks of thrombosis and bleeding. The newer target-specific oral anticoagulants (dabigatran, apixaban, rivaroxaban, and edoxaban) are in increasingly common use. Simple extractions of one to three teeth, periodontal surgery, abscess drainage, and implant positioning do not typically require interruption of therapy. More extensive surgery may necessitate delaying or holding a dose of the anticoagulant or more elaborate measures to manage the risk of thrombosis and bleeding.

Patients at risk for bacterial endocarditis (**Chap. 128**) should maintain optimal oral hygiene, including flossing, and have regular professional cleanings. Currently, guidelines recommend that prophylactic antibiotics be restricted to those patients at high risk for bacterial endocarditis who undergo dental and oral procedures involving significant manipulation of gingival or periapical tissue or penetration of the oral mucosa. If unexpected bleeding occurs, antibiotics given within 2 h after the procedure provide effective prophylaxis.

Hematogenous bacterial seeding from oral infection can undoubtedly produce late prosthetic-joint infection and therefore requires removal of the infected tissue (e.g., drainage, extraction, root canal) and appropriate antibiotic therapy. However, evidence that late prosthetic-joint infection follows routine dental procedures is lacking. For this reason, antibiotic prophylaxis is generally not recommended before oral surgery or oral mucosal manipulation for patients who have undergone joint replacement surgery. Exceptions to this may be considered for patients who have experienced joint replacement complications.

Concern often arises regarding the use of vasoconstrictors to treat patients with hypertension and heart disease. Vasoconstrictors enhance the depth and duration of local anesthesia, thus reducing the anesthetic dose and potential toxicity. If intravascular injection is avoided, 2% lidocaine with 1:100,000 epinephrine (limited to a total of 0.036 mg of epinephrine) can be used safely in patients with controlled hypertension and stable coronary heart disease, arrhythmia,

or congestive heart failure. Precautions should be taken with patients taking tricyclic antidepressants and nonselective beta blockers because these drugs may potentiate the effect of epinephrine.

Elective dental treatments should be postponed for at least 1 month and preferably for 6 months after myocardial infarction, after which the risk of reinfarction is low provided the patient is medically stable (e.g., stable rhythm, stable angina, and no heart failure). Patients who have suffered a stroke should have elective dental care deferred for 9 months. In both situations, effective stress reduction requires good pain control, including the use of the minimal amount of vasoconstrictor necessary to provide good hemostasis and local anesthesia.

Bisphosphonate therapy is associated with *osteonecrosis* of the jaw. However, the risk with oral bisphosphonate therapy is very low. Most patients affected have received high-dose aminobisphosphonate therapy for multiple myeloma or metastatic breast cancer and have undergone tooth extraction or dental surgery. Intraoral lesions, of which two-thirds are painful, appear as exposed yellow-white hard bone involving the mandible or maxilla. Screening tests for determining risk of osteonecrosis are unreliable. Patients slated for aminobisphosphonate therapy should receive preventive dental care that reduces the risk of infection and the need for future dentoalveolar surgery.

Halitosis Halitosis typically emanates from the oral cavity or nasal passages. Volatile sulfur compounds resulting from bacterial decay of food and cellular debris account for the malodor. Periodontal disease, caries, acute forms of gingivitis, poorly fitting dentures, oral abscess, and tongue coating are common causes. Treatment includes correcting poor hygiene, treating infection, and tongue brushing. Hyposalivation can produce and exacerbate halitosis. Pockets of decay in the tonsillar crypts, esophageal diverticulum, esophageal stasis (e.g., achalasia, stricture), sinusitis, and lung abscess account for some instances. A few systemic diseases produce distinctive odors: renal failure (ammoniacal), hepatic (fishy), and ketoacidosis (fruity). *Helicobacter pylori* gastritis can also produce ammoniacal breath. If a patient presents because of concern about halitosis but no odor is detectable, then pseudohalitosis or halitophobia must be considered.

Aging and Oral Health While tooth loss and dental disease are not normal consequences of aging, a complex array of structural and functional changes that occur with age can affect oral health. Subtle changes in tooth structure (e.g., diminished pulp space and volume, sclerosis of dentinal tubules, and altered proportions of nerve and vascular pulp content) result in the elimination or diminution of pain sensitivity and a reduction in the reparative capacity of the teeth. In addition, age-associated fatty replacement of salivary acini may reduce physiologic reserve, thus increasing the risk of hyposalivation. In healthy older adults, there is minimal, if any, reduction in salivary flow.

Poor oral hygiene often results when general health fails or when patients lose manual dexterity and upper-extremity flexibility. This situation is particularly common among frail older adults and nursing home residents and must be emphasized because regular oral cleaning and dental care reduce the incidence of pneumonia and oral disease as well as the mortality risk in this population. Other risks for dental decay include limited lifetime fluoride exposure. Without assiduous care, decay can become quite advanced yet remain asymptomatic. Consequently, much of a tooth—or the entire tooth—can be destroyed before the patient is aware of the process.

Periodontal disease, a leading cause of tooth loss, is indicated by loss of alveolar bone height. More than 90% of the U.S. population has some degree of periodontal disease by age 50. Healthy adults who have not had significant alveolar bone loss by the sixth decade of life do not typically experience significant worsening with advancing age.

With the passing of those born in the first half of the twentieth century, complete edentulousness in the United States is becoming increasingly restricted to impoverished populations. When it is present, speech, mastication, and facial contours are dramatically affected. Edentulousness may also exacerbate obstructive sleep apnea, particularly in asymptomatic individuals who wear dentures. Dentures can

improve verbal articulation and restore diminished facial contours. Mastication can also be restored; however, patients expecting dentures to facilitate oral intake are often disappointed. Accommodation to dentures requires a period of adjustment. Pain can result from friction or traumatic lesions produced by loose dentures. Poor fit and poor oral hygiene may permit the development of candidiasis. This fungal infection may be either asymptomatic or painful and is suggested by erythematous smooth or granular tissue conforming to an area covered by the appliance. Individuals with dentures and no natural teeth need regular (annual) professional oral examinations.

FURTHER READING

- DURSO SC: Interaction with other health team members in caring for elderly patients. *Dent Clin North Am* 49:377, 2005.
- KAPLOVITCH E, DOUNAEVSKAIA V: Treatment in the dental practice of the patient receiving anticoagulant therapy. *J Am Dent Assoc* 150:602, 2019.
- WEINTRAUB JA et al: Improving nursing home residents' oral hygiene: Results of a cluster randomized intervention trial. *J Am Med Dir Assoc* 19:1086, 2018.

MECHANISMS UNDERLYING DYSPNEA

The mechanisms underlying dyspnea are complex, as it can arise from different contributory respiratory sensations. Although a large body of research has increased our understanding of mechanisms underlying particular respiratory sensations such as "chest tightness" or "air hunger," it is likely that a given disease state might produce the sensation of dyspnea via more than one underlying mechanism. Dyspnea can arise from a variety of pathways, including generation of *afferent* signals from the respiratory system to the central nervous system (CNS), *efferent* signals from the CNS to the respiratory muscles, and particularly when there is a mismatch in the integrative signaling between these two pathways, termed *efferent-reafferent mismatch* (**Fig. 37-1**).

Afferent signals trigger the CNS (brainstem and/or cortex) and include primarily: (1) peripheral chemoreceptors in the carotid body and aortic arch and central chemoreceptors in the medulla that are activated by hypoxemia, hypercapnia, or acidemia, and might produce a sense of "air hunger"; and (2) mechanoreceptors in the upper airways, lungs (including stretch receptors, irritant receptors, and J receptors), and chest wall (including muscle spindles as stretch receptors and tendon organs that monitor force generation) that are activated in the setting of an increased work load from a disease state producing an increase in airway resistance that may be associated with symptoms of chest tightness (e.g., asthma or COPD) or decreased lung or chest wall compliance (e.g., pulmonary fibrosis). Other afferent signals that trigger dyspnea within the respiratory system can arise from pulmonary vascular receptor responses to changes in pulmonary artery pressure and skeletal muscle (termed metaboreceptors) that are believed to sense changes in the biochemical environment.

Efferent signals are sent from the CNS (motor cortex and brainstem) to the respiratory muscles and are also transmitted by corollary discharge to the sensory cortex; they are believed to underlie sensations of respiratory effort (or "work of breathing") and perhaps contribute to sensations of "air hunger," especially in response to an increased ventilatory load in a disease state such as COPD. In addition, fear or anxiety may heighten the sense of dyspnea by exacerbating the underlying physiologic disturbance in response to an increased respiratory rate or disordered breathing pattern.

Section 5 Alterations in Circulatory and Respiratory Functions

37

Dyspnea

Rebecca M. Baron



DYSPNEA

DEFINITION

The American Thoracic Society consensus statement defines *dyspnea* as a "subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity. The experience derives from interactions among multiple physiological, psychological, social, and environmental factors and may induce secondary physiological and behavioral responses." Dyspnea, a symptom, can be perceived only by the person experiencing it and, therefore, must be self-reported. In contrast, signs of increased work of breathing, such as tachypnea, accessory muscle use, and intercostal retraction, can be measured and reported by clinicians.

EPIDEMIOLOGY

Dyspnea is common. It has been reported that up to one-half of inpatients and one-quarter of ambulatory patients experience dyspnea, with a prevalence of 9–13% in the community that increases to as high as 37% for adults aged ≥70 years. Dyspnea is a frequent cause for emergency room visits, accounting for as many as 3–4 million visits per year. Furthermore, it is increasingly appreciated that the degree of dyspnea may better predict outcomes in chronic obstructive pulmonary disease (COPD) than does the forced expiratory volume in 1 s (FEV₁), and formal measures of dyspnea have been incorporated into the Global Initiative for Chronic Obstructive Lung Disease (GOLD) COPD severity assessment guidelines. Dyspnea may also predict outcomes in other chronic heart and lung diseases as well. Dyspnea can arise from a diverse array of pulmonary, cardiac, and neurologic underlying causes, and elucidation of particular symptoms may point toward a specific etiology and/or mechanism driving dyspnea (although additional diagnostic testing is often required, as will be further discussed below).

ASSESSING DYSPNEA

While it is well appreciated that dyspnea is a difficult quality to reliably measure due to multiple relevant possible domains that can be measured (e.g., sensory-perceptual experience, affective distress, and symptom impact or burden), and there exist no uniformly agreed upon tools for dyspnea assessment, consensus opinion is that dyspnea should be formally assessed in a context most relevant and beneficial for patient management and, furthermore, that the specific domains being measured are adequately described. There are a number of emerging tools that have been developed for formal dyspnea assessment. As an example, the GOLD criteria advocate use of a dyspnea assessment tool such as the Modified Medical Research Council Dyspnea Scale (**Table 37-1**) to assess symptom/impact burden in COPD.

DIFFERENTIAL DIAGNOSIS

This chapter focuses largely on chronic dyspnea, which is defined as symptoms lasting longer than 1 month and can arise from a broad array of different underlying conditions, most commonly attributable to pulmonary or cardiac conditions that account for as many as 85% of the underlying causes of dyspnea. However, as many as one-third of patients may have multifactorial reasons underlying dyspnea. Examples of a wide array of conditions that underlie dyspnea with possible mechanisms underlying the presenting symptoms are described in **Table 37-2**.

Respiratory system causes include diseases of the airways (e.g., asthma and COPD), diseases of the parenchyma (more commonly, interstitial lung diseases are seen in the setting of chronic dyspnea, but alveolar filling processes, such as hypersensitivity pneumonitis or bronchiolitis obliterans organizing pneumonia [BOOP], can also

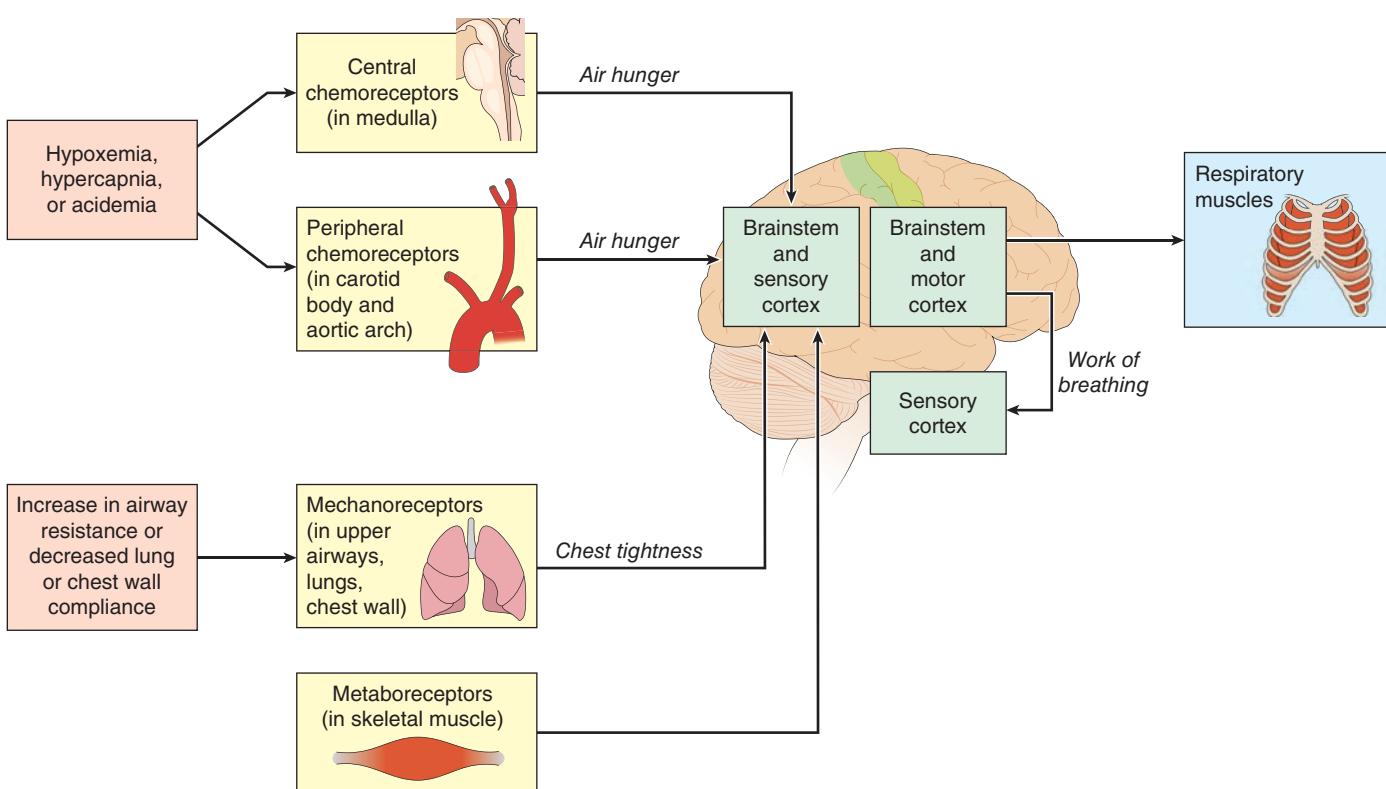


FIGURE 37-1 Signaling pathways underlying dyspnea. Dyspnea arises from a range of sensory inputs, many of which lead to distinct descriptive phrases used by patients (shown in italics in the figure). The sensation of respiratory effort (or work of breathing) likely arises from signals transmitted from the motor cortex to the sensory cortex when outgoing motor commands are sent to the respiratory muscles. Motor output from the brain stem may also be accompanied by signals transmitted to the sensory cortex and contribute to the sensation of work of breathing. The sensation of air hunger likely derives from stimuli that increase the drive to breathe (e.g., hypoxemia, hypercapnia, acidemia; mediated by signals from central and peripheral chemoreceptors), as well as airway and interstitial inflammation (mediated by pulmonary afferent signals) and pulmonary vascular receptors. Dyspnea arises, in part, from a perceived mismatch between the outgoing efferent messages to the respiratory muscles and incoming afferent signals from the lungs and chest wall. Chest tightness, often associated with bronchospasm, is largely mediated by simulation of vagal-irritant receptors. Afferent signals from airway, lung, and chest wall mechanoreceptors most likely pass through the brain stem before being transmitted to the sensory cortex, although it is possible that some afferent information bypasses the brain stem and goes directly to the sensory cortex. (Adapted from RM Schwartzstein: Approach to the patient with dyspnea. In: UpToDate, TW Post (Ed), UpToDate, Waltham, MA. (Accessed on 7 December 2021) 2018 UpToDate, Inc. For more information visit www.uptodate.com.)

present with similar symptoms), diseases affecting the chest wall (e.g., bony abnormalities such as kyphoscoliosis, or neuromuscular weakness conditions such as amyotrophic lateral sclerosis), and diseases affecting the pulmonary vasculature (e.g., pulmonary hypertension that can

arise from a variety of underlying causes, or chronic thromboembolic disease). Diseases affecting the cardiovascular system that can present with dyspnea include processes affecting left heart function, such as coronary artery disease and cardiomyopathy, as well as disease processes affecting the pericardium, including constrictive pericarditis and cardiac tamponade. Other conditions underlying dyspnea that might not directly emanate from the pulmonary or cardiovascular systems include anemia (thereby potentially affecting oxygen-carrying capacity), deconditioning, and psychological processes such as anxiety. Distinguishing between the myriad of underlying processes that might present with dyspnea can be challenging. A graded approach that begins with a history and physical examination, followed by selected laboratory testing that might then advance to additional diagnostics and potentially subspecialty referral, may help elucidate the underlying cause of dyspnea. However, a substantial proportion of patients may have persistent dyspnea despite treatment for an underlying process or may not have a specific underlying process identified that is driving the dyspnea.

TABLE 37-1 An Example of a Clinical Method for Rating Dyspnea: The Modified Medical Research Council Dyspnea Scale*

GRADE OF DYSPNEA	DESCRIPTION
0	Not troubled by breathlessness, except with strenuous exercise
1	Shortness of breath walking on level ground or with walking up a slight hill
2	Walks slower than people of similar age on level ground due to breathlessness, or has to stop to rest when walking at own pace on level ground
3	Stops to rest after walking 100 m or after walking a few minutes on level ground
4	Too breathless to leave the house, or breathless with activities of daily living (e.g., dressing/undressing)

*Which has been incorporated into the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines as a possible tool for rating dyspnea in chronic obstructive pulmonary disease.

Source: Reproduced with permission from DA Mahler, CK Wells: Evaluation of clinical methods for rating dyspnea. *Chest* 93:580, 1988.

APPROACH TO THE PATIENT

Dyspnea (See Fig. 37-2)

OVERALL

For patients with a known prior pulmonary, cardiac, or neuromuscular condition and worsening dyspnea, the initial focus of the

TABLE 37-2 Differential Diagnosis of Disease Processes Underlying Dyspnea

SYSTEM	TYPE OF PROCESS	EXAMPLE OF DISEASE PROCESS	POSSIBLE PRESENTING DYSPNEA SYMPTOMS	POSSIBLE PHYSICAL FINDINGS	POSSIBLE MECHANISMS UNDERLYING DYSPNEA	INITIAL DIAGNOSTIC STUDIES (AND POSSIBLE FINDINGS)
Pulmonary	Airways disease	Asthma, COPD, upper airway obstruction	Chest tightness, tachypnea, increased WOB, air hunger, inability to get a deep breath	Wheezing, accessory muscle use, exertional hypoxemia (especially with COPD)	Increased WOB, hypoxemia, hypercapnia, stimulation of pulmonary receptors	Peak flow (reduced); spirometry (OVD); CXR (hyperinflation; loss of lung parenchyma in COPD), chest CT and airway examination for upper airway obstruction
	Parenchymal disease	Interstitial lung disease ^a	Air hunger, inability to get a deep breath	Dry end-inspiratory crackles, clubbing, exertional hypoxemia	Increased WOB, increased respiratory drive, hypoxemia, hypercapnia, stimulation of pulmonary receptors	Spirometry and lung volumes (RVD); CXR and chest CT (interstitial lung disease)
	Chest wall disease	Kyphoscoliosis, neuromuscular (NM) weakness	Increased WOB, inability to get a deep breath	Decreased diaphragm excursion; atelectasis	Increased WOB; stimulation of pulmonary receptors (if atelectasis is present)	Spirometry and lung volumes (RVD); MIP and MEPs (reduced in NM weakness)
Pulmonary and cardiac	Pulmonary vasculature	Pulmonary hypertension	Tachypnea	Elevated right heart pressures, exertional hypoxemia	Increased respiratory drive, hypoxemia, stimulation of vascular receptors	Diffusion capacity (reduced); ECG; ECHO (to evaluate pulmonary artery pressures) ^b
Cardiac	Left heart failure	Coronary artery disease, cardiomyopathy ^c	Chest tightness, air hunger	Elevated left heart pressures; wet crackles on lung examination; pulsus paradoxus (pericardial disease)	Increased WOB and drive, hypoxemia, stimulation of vascular and pulmonary receptors ^d	Consider BNP testing, especially in the acute setting; ECG, ECHO, may need stress testing and/or LHC
	Pericardial disease	Constrictive pericarditis; cardiac tamponade				
Other	Variable	Anemia Deconditioning Psychological Metabolic disturbances Gastrointestinal (e.g., gastroesophageal reflux disease [GERD], aspiration pneumonitis)	Exertional breathlessness Poor fitness Anxiety	Variable	Metaboreceptors (anemia, poor fitness); chemoreceptors (anaerobic metabolism from poor fitness); some subjects may have increased sensitivity to hypercapnia	Hematocrit for anemia; laboratory studies (e.g., metabolic panel, thyroid hormone testing for metabolic disturbances); consider upper gastrointestinal endoscopy and/or esophageal pH probe testing for GERD and concerns for aspiration; exclude other causes

^aDifferential diagnosis of interstitial lung disease includes idiopathic pulmonary fibrosis, collagen vascular disease, drug- or occupation-induced pneumonitis, lymphangitic spread of malignancy; processes that are more alveolar rather than interstitial in nature can also less commonly contribute to parenchymal lung disease underlying chronic dyspnea and include entities such as hypersensitivity pneumonitis, bronchiolitis obliterans organizing pneumonia, etc. ^bWould additionally consider these patients for CT angiography to evaluate for presence of thromboemboli, ventilation/perfusion scanning to evaluate for the presence of chronic thromboembolic disease, and right heart catheterization to further evaluate for pulmonary hypertension. ^cDiastolic dysfunction in the setting of a stiff left ventricle is often seen and contributes significantly to insidious dyspnea that can be difficult to treat. ^dMay stimulate metaboreceptors if cardiac output is sufficiently reduced to result in a lactic acidosis.

Abbreviations: BNP, brain natriuretic peptide; COPD, chronic obstructive pulmonary disease; CT, computed tomography; CXR, chest x-ray; ECG, electrocardiogram; ECHO, echocardiogram; GERD, gastroesophageal reflux disease; LHC, left heart catheterization; MIP/MEP, maximal inspiratory and maximal expiratory pressures (obtained in the pulmonary function testing laboratory); OVD, obstructive ventilatory defect; RVD, restrictive ventilatory defect; WOB, work of breathing.

evaluation will usually address determining whether the known condition has progressed or whether a new process has developed that is causing dyspnea. For patients without a prior known potential cause of dyspnea, the initial evaluation will focus on determining an underlying etiology. Determining the underlying cause, if possible, is extremely important, as the treatment may vary dramatically based on the predisposing condition. An initial history and physical examination remain fundamental to the evaluation followed by initial diagnostic testing as indicated that might prompt subspecialty referral (e.g., pulmonary, cardiology, neurology, sleep, and/or specialized dyspnea clinic) if the cause of dyspnea remains elusive (Fig. 37-2). As many as two-thirds of patients will require diagnostic testing beyond the initial clinical presentation.

HISTORY

The patient should be asked to describe in his or her own words what the discomfort feels like as well as the effect of position,

infections, and environmental stimuli on the dyspnea, as descriptors may be helpful in pointing toward an etiology. For example, symptoms of chest tightness might suggest the possibility of bronchoconstriction, and the sensation of inability to take a deep breath may correlate with dynamic hyperinflation from COPD. Orthopnea is a common indicator of congestive heart failure (CHF), mechanical impairment of the diaphragm associated with obesity, or asthma triggered by esophageal reflux. Nocturnal dyspnea suggests CHF or asthma. Acute, intermittent episodes of dyspnea are more likely to reflect episodes of myocardial ischemia, bronchospasm, or pulmonary embolism, while chronic persistent dyspnea is more typical of COPD, interstitial lung disease, and chronic thromboembolic disease. Information on risk factors for drug-induced or occupational lung disease and for coronary artery disease should be elicited. Left atrial myxoma or hepatopulmonary syndrome should be considered when the patient complains of *platypnea*—i.e., dyspnea in the upright position with relief in the supine position.

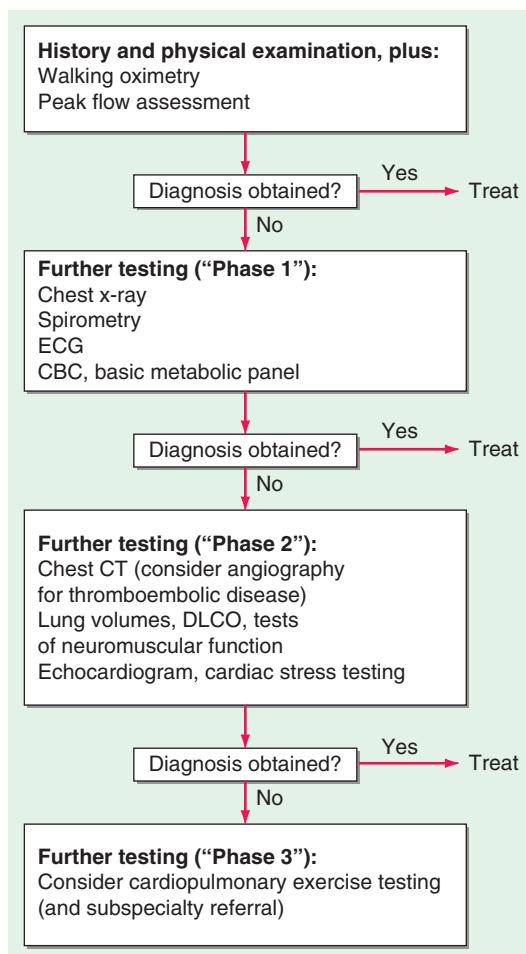


FIGURE 37-2 Possible algorithm for the evaluation of the patient with dyspnea. As described in the text, the approach should begin with a detailed history and physical examination, followed by progressive testing and ultimately more invasive testing and subspecialty referral as is indicated to determine the underlying cause of dyspnea. CBC, complete blood count; DLCO, diffusing capacity of the lungs for carbon monoxide; ECG, electrocardiogram. (Adapted from NG Karnani et al: Am Fam Physician 71:1529, 2005.)

PHYSICAL EXAMINATION

Initial vital signs might be helpful in pointing toward an underlying etiology in the context of the remainder of the evaluation. For example, the presence of fever might point toward an underlying infectious or inflammatory process; the presence of hypertension in the setting of a heart failure might point toward diastolic dysfunction; the presence of tachycardia might be associated with many different underlying processes including fever, cardiac dysfunction, and deconditioning; and the presence of resting hypoxemia suggests processes involving hypercapnia, ventilation-perfusion mismatch, shunt, or impairment in diffusion capacity might be involved. An exertional oxygen saturation should also be obtained as described below. The physical examination should begin during the interview of the patient. Inability of the patient to speak in full sentences before stopping to get a deep breath suggests a condition that leads to stimulation of the controller or impairment of the ventilatory pump with reduced vital capacity. Evidence of increased work of breathing (supraclavicular retractions; use of accessory muscles of ventilation; and the tripod position, characterized by sitting with the hands braced on the knees) is indicative of increased airway resistance or stiffness of the lungs and the chest wall. When measuring the vital signs, the physician should accurately assess the respiratory rate and measure the pulsus paradoxus (**Chap. 270**); if the systolic pressure decreases by >10 mmHg on inspiration, the

presence of COPD, acute asthma, or pericardial disease should be considered. During the general examination, signs of anemia (pale conjunctivae), cyanosis, and cirrhosis (spider angiomas, gynecomastia) should be sought. Examination of the chest should focus on symmetry of movement; percussion (dullness is indicative of pleural effusion; hyperresonance is a sign of pneumothorax and emphysema); and auscultation (wheezes, rhonchi, prolonged expiratory phase, and diminished breath sounds are clues to disorders of the airways; rales suggest interstitial edema or fibrosis). The cardiac examination should focus on signs of elevated right heart pressures (jugular venous distention, edema, accentuated pulmonic component to the second heart sound); left ventricular dysfunction (S3 and S4 gallops); and valvular disease (murmurs). When examining the abdomen with the patient in the supine position, the physician should note whether there is paradoxical movement of the abdomen as well as the presence of increased respiratory distress in the supine position: inward motion during inspiration is a sign of diaphragmatic weakness, and rounding of the abdomen during exhalation is suggestive of pulmonary edema. Clubbing of the digits may be an indication of interstitial pulmonary fibrosis or bronchiectasis, and joint swelling or deformation as well as changes consistent with Raynaud's disease may be indicative of a collagen-vascular process that can be associated with pulmonary disease.

Patients should be asked to walk under observation with oximetry in order to reproduce the symptoms. The patient should be examined during and at the end of exercise for new findings that were not present at rest (e.g., presence of wheezing) and for changes in oxygen saturation.

CHEST IMAGING

After the history elicitation and the physical examination, a chest radiograph should be obtained if the diagnosis remains elusive. The lung volumes should be assessed: hyperinflation is consistent with obstructive lung disease, whereas low lung volumes suggest interstitial edema or fibrosis, diaphragmatic dysfunction, or impaired chest wall motion. The pulmonary parenchyma should be examined for evidence of interstitial disease, infiltrates, and emphysema. Prominent pulmonary vasculature in the upper zones indicates pulmonary venous hypertension, while enlarged central pulmonary arteries may suggest pulmonary arterial hypertension. An enlarged cardiac silhouette can point toward dilated cardiomyopathy or valvular disease. Bilateral pleural effusions are typical of CHF and some forms of collagen-vascular disease. Unilateral effusions raise the specter of carcinoma and pulmonary embolism but may also occur in heart failure or in the case of a parapneumonic effusion. CT of the chest is generally reserved for further evaluation of the lung parenchyma (interstitial lung disease) and possible pulmonary embolism if there remains diagnostic uncertainty.

LABORATORY STUDIES

Initial laboratory testing should include a hematocrit to exclude occult anemia as an underlying cause of reduced oxygen-carrying capacity contributing to dyspnea, and a basic metabolic panel may be helpful to exclude a significant underlying metabolic acidosis (and conversely, an elevated bicarbonate might point toward the possibility of carbon dioxide retention that might be seen in chronic respiratory failure—in such a setting, an arterial blood gas may provide useful additional information). Additional laboratory studies should include electrocardiography to seek evidence of ventricular hypertrophy and prior myocardial infarction and spirometry, which can be diagnostic of the presence of an obstructive ventilatory defect and suggest the possibility of a restrictive ventilatory defect (that then might prompt additional pulmonary function laboratory testing, including lung volumes, diffusion capacity, and possible tests of neuromuscular function). Echocardiography is indicated when systolic dysfunction, pulmonary hypertension, or

valvular heart disease is suspected. Bronchoprovocation testing and/or home peak-flow monitoring may be useful in patients with intermittent symptoms suggestive of asthma who have a normal physical examination and spirometry; up to one-third of patients with the clinical diagnosis of asthma do not have reactive airways disease when formally tested. Measurement of brain natriuretic peptide levels in serum is increasingly used to assess for CHF in patients presenting with acute dyspnea but may be elevated in the presence of right ventricular strain as well.

DISTINGUISHING CARDIOVASCULAR FROM RESPIRATORY SYSTEM DYSPNEA

If a patient has evidence of both pulmonary and cardiac disease that is not responsive to treatment or it remains unclear what factors are primarily driving the dyspnea, a cardiopulmonary exercise test (CPET) can be carried out to determine which system is responsible for the exercise limitation. CPET includes incremental symptom-limited exercise (cycling or treadmill) with measurements of ventilation and pulmonary gas exchange and, in some cases, includes noninvasive and invasive measures of pulmonary vascular pressures and cardiac output. If, at peak exercise, the patient achieves predicted maximal ventilation, demonstrates an increase in dead space or hypoxemia, or develops bronchospasm, the respiratory system may be the cause of the problem. Alternatively, if the heart rate is >85% of the predicted maximum, if the anaerobic threshold occurs early, if the blood pressure becomes excessively high or decreases during exercise, if the O₂ pulse (O₂ consumption/heart rate, an indicator of stroke volume) falls, or if there are ischemic changes on the electrocardiogram, an abnormality of the cardiovascular system is likely the explanation for the breathing discomfort. Additionally, a CPET may also help point toward a peripheral extraction deficit or metabolic/neuromuscular disease as potential underlying processes driving dyspnea.

TREATMENT

Dyspnea

The first goal is to correct the underlying condition(s) driving dyspnea and address potentially reversible causes with appropriate treatment for the particular condition. Multiple different interventions may be necessary, given that dyspnea often arises from multifactorial causes. If relief of dyspnea with treatment of the underlying condition(s) is not fully possible, an effort is made to lessen the intensity of the symptom and its effect on the patient's quality of life. More recent work at the consensus conference level has sought to define an identifiable entity of persistent dyspnea in order to develop an approach to improving efforts to address symptom management for this condition. In 2017, an international group of experts defined "chronic breathlessness syndrome" as "the experience of breathlessness that persists despite optimal treatment of the underlying pathophysiology and results in disability for the patient." Despite an increased understanding of the mechanisms underlying dyspnea, there has been limited progress in treatment strategies for dyspnea. Supplemental O₂ should be administered if the resting O₂ saturation is ≤88% or if the patient's saturation drops to these levels with activity or sleep. In particular, for patients with COPD, supplemental oxygen for those with hypoxemia has been shown to improve mortality, and pulmonary rehabilitation programs (including some community-based exercise programs such as yoga and Tai Chi) have demonstrated positive effects on dyspnea, exercise capacity, and rates of hospitalization. Opioids have been shown to reduce symptoms of dyspnea, largely through reducing air hunger, thus likely suppressing respiratory drive and influencing cortical activity. However, opioids should be considered for each patient individually based on the risk-benefit profile in regard to the effects of respiratory depression. Studies of anxiolytics for dyspnea

have not demonstrated consistent benefit. Additional approaches are under study for dyspnea, including inhaled furosemide that might alter afferent sensory information.

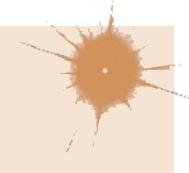
FURTHER READING

- BANZETT RB et al: Multidimensional dyspnea profile: An instrument for clinical and laboratory research. Eur Respir J 45:1681, 2015.
- FERRY OR et al: Diagnostic approach to chronic dyspnea in adults. J Thorac Dis 11(Suppl 17):S2117, 2019.
- JOHNSON M et al: Toward an expert consensus to delineate a clinical syndrome of chronic breathlessness: Chronic breathlessness syndrome. Eur Respir J 49:1602277, 2017.
- LAVIOLETTE L, LAVENEZIANA P ON BEHALF OF THE ERS RESEARCH SEMINAR FACULTY: Dyspnoea: A multidimensional and multidisciplinary approach. Eur Respir J 43:1750, 2014.
- O'DONNELL DE et al: Unraveling the causes of unexplained dyspnea. Clin Chest Med 40:471, 2019.
- PARSHALL MB et al: An Official American Thoracic Society Statement: Update on the mechanisms, assessment, and management of dyspnea. Am J Respir Crit Care Med 185:435, 2012.
- RATARASARN K et al: Yoga and Tai Chi: A mind-body approach in managing respiratory symptoms in obstructive lung diseases. Curr Opin Pulm Med 26:186, 2020.

38

Cough

Christopher H. Fanta



COUGH

Cough performs an essential protective function for human airways and lungs. Without an effective cough reflex, we are at risk for retained airway secretions and aspirated material predisposing to infection, atelectasis, and respiratory compromise. At the other extreme, excessive coughing can be exhausting; can be complicated by emesis, syncope, muscular pain, or rib fractures; can aggravate low back pain, abdominal or inguinal hernias, and urinary incontinence; and can be a major impediment to social interactions. Cough is often a clue to the presence of respiratory disease. In many instances, cough is an expected and accepted manifestation of disease, as in acute respiratory tract infection. However, persistent cough in the absence of other respiratory symptoms commonly causes patients to seek medical attention.

COUGH MECHANISM

Both chemical (e.g., capsaicin) and mechanical (e.g., mucus, particulates in air pollution) stimuli can initiate the cough reflex. Cationic channels (e.g., transient receptor potential channels) and adenosine triphosphate-activated ion channels (P2X3) function as sensory neuronal receptors, with signals transmitted centrally via A_δ (mechanosensory) and C fibers (chemosensory). Afferent nerve endings richly innervate the pharynx, larynx, and airways to the level of the terminal bronchioles and extend into the lung parenchyma. They are also located in the external auditory canal (the auricular branch of the vagus nerve, or Arnold's nerve) and in the esophagus. Sensory signals travel via the vagus and superior laryngeal nerves to a region of the brainstem in the nucleus tractus solitarius. Integrated neural networks process this input into a conscious sensation referred to as the "urge to cough." The efferent limb of the cough reflex involves a highly orchestrated series of involuntary muscular actions, with the potential for input from cortical pathways as well, making possible voluntary cough. The vocal

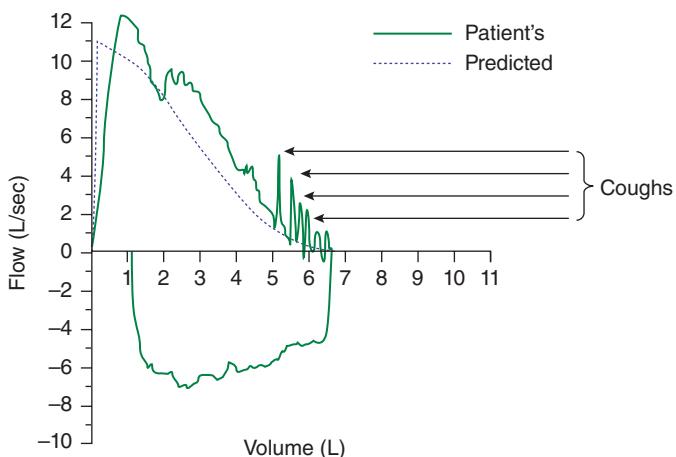


FIGURE 38-1 Flow-volume curve shows spikes of high expiratory flow achieved with cough.

cords adduct, leading to transient upper-airway occlusion. Expiratory muscles contract, generating positive intrathoracic pressures as high as 300 mmHg. With sudden release of the laryngeal contraction, rapid expiratory flows are generated, exceeding the normal “envelope” of maximal expiratory flow seen on the flow-volume curve (Fig. 38-1). Bronchial smooth-muscle contraction together with dynamic compression of airways narrows airway lumens and maximizes the velocity of exhalation. The kinetic energy available to dislodge mucus from the inside of airway walls is directly proportional to the square of the velocity of expiratory airflow. A deep breath preceding a cough optimizes the function of the expiratory muscles; a series of repetitive coughs at successively lower lung volumes sweeps the point of maximal expiratory velocity progressively further into the lung periphery.

■ IMPAIRED COUGH

Weak or ineffective cough compromises the ability to clear lower respiratory tract secretions, predisposing to more serious infections and their sequelae. Weakness or paralysis of the expiratory (abdominal and intercostal) muscles and pain in the chest wall or abdomen are foremost on the list of causes of impaired cough (Table 38-1). Cough strength is generally assessed qualitatively; peak expiratory flow or maximal expiratory pressure at the mouth can be used as a surrogate marker for cough strength. A variety of assistive devices and techniques have been developed to improve cough efficacy, running the gamut from simple (splinting of the abdominal muscles with a tightly held pillow to reduce postoperative pain while coughing) to complex (a mechanical cough-assist device supplied via face mask or tracheal tube that applies a cycle of positive pressure followed rapidly by negative pressure). Cough may fail to clear secretions completely despite a preserved ability to generate normal expiratory velocities; such failure may be due to abnormal airway secretions (e.g., abnormally viscous secretions of cystic fibrosis), ciliary dysfunction (e.g., primary ciliary dyskinesia), or structural abnormalities of the airways (e.g., tracheomalacia with excessive expiratory collapse of the trachea during cough).

TABLE 38-1 Causes of Impaired Cough and Airway Clearance

Respiratory muscle weakness
Chest wall or abdominal pain
Chest wall deformity (e.g., severe kyphoscoliosis)
Impaired glottic closure or tracheostomy
Central respiratory depression (e.g., anesthesia, sedation, or neurologic disease)
Abnormal airway secretions
Ciliary dysfunction
Tracheobronchomalacia
Bronchiectasis
Tracheal or bronchial stenoses

■ SYMPTOMATIC COUGH

Cough may occur in the context of other respiratory symptoms that together point to a diagnosis; for example, cough accompanied by wheezing, shortness of breath, and chest tightness after exposure to a cat or other sources of allergens suggests asthma. At times, however, cough is the dominant or sole symptom of disease, and it may be of sufficient duration and severity that relief is sought. The duration of cough is a clue to its etiology, at least retrospectively. Acute cough (<3 weeks) is most commonly due to a respiratory tract infection, aspiration, or inhalation of noxious chemicals or smoke. Subacute cough (3–8 weeks in duration) is a common residuum of tracheobronchitis, as in pertussis or “postviral tussive syndrome.” Chronic cough (>8 weeks) may be caused by a wide variety of cardiopulmonary diseases, including those of inflammatory, infectious, neoplastic, and cardiovascular etiologies. When initial assessment with chest examination and radiography is normal, cough-variant asthma, gastroesophageal reflux, rhinosinusitis with excessive nasopharyngeal drainage, and medications (angiotensin-converting enzyme [ACE] inhibitors) are the most common identifiable causes of chronic cough. In a long-time cigarette smoker, an early-morning, productive cough suggests chronic bronchitis. A dry, irritative cough that lingers for >2 months following one or more respiratory tract infections (“postbronchitic cough”) is a very common cause of chronic cough, especially in the winter months. Chronic cough in the absence of identifiable etiology has been recognized with increasing frequency, is thought to be due to exaggerated neurologic signaling via sensory cough-reflex pathways, and is referred to as “chronic cough hypersensitivity syndrome.”

■ ASSESSMENT OF CHRONIC COUGH

Except for our ability to detect the sound of excess airway secretions, details as to the resonance of the cough, its time of occurrence during the day, and the pattern of coughing (e.g., occurring in paroxysms) infrequently provide useful etiologic clues. Regardless of cause, cough often worsens upon first lying down at night, with talking, or with the hyperpnea of exercise; it frequently improves with sleep. An exception may involve the cough that occurs only with certain allergic exposures or exercise in cold air, as in asthma. Useful historical questions include what circumstances surrounded the onset of cough, what makes the cough better or worse, and whether the cough produces sputum.

The physical examination seeks clues suggesting the presence of cardiopulmonary disease, including findings such as wheezing or crackles on chest examination. Examination of the auditory canals and tympanic membranes (for irritation of the latter resulting in stimulation of Arnold’s nerve), the nasal passageways (for rhinitis or polyps), and the nails (for clubbing) may also provide etiologic clues. Because cough can be a manifestation of a systemic disease such as sarcoidosis or vasculitis, a thorough general examination is likewise important.

In virtually all instances, evaluation of chronic cough merits a chest radiograph. The list of diseases that can cause persistent cough without other symptoms and without detectable abnormalities on physical examination is long. It includes serious illnesses such as sarcoidosis or Hodgkin’s disease in young adults, lung cancer in older patients, and (worldwide) pulmonary tuberculosis. An abnormal chest film prompts an evaluation aimed at explaining the radiographic abnormality. In a patient with chronic productive cough, examination of expectorated sputum is warranted, because determining the cause of mucus hypersecretion is a crucial clue to etiology. Purulent-appearing sputum should be sent for routine bacterial culture and, in certain circumstances, mycobacterial culture as well. Cytologic examination of mucoid sputum may be useful to assess for malignancy and oropharyngeal aspiration and to distinguish neutrophilic from eosinophilic bronchitis. Expectoration of blood—whether streaks of blood, blood mixed with airway secretions, or pure blood—deserves a special approach to assessment and management (Chap. 39).

■ CHRONIC COUGH WITH A NORMAL CHEST RADIOGRAPH

It is commonly held that (alone or in combination) the use of an ACE inhibitor; postnasal drainage; gastroesophageal reflux; and asthma

account for >90% of cases of chronic cough with a normal or noncontributory chest radiograph. However, clinical experience does not support this contention, and strict adherence to this concept discourages the search for alternative explanations by both clinicians and researchers. In recent years, the concept of a distinct “cough hypersensitivity syndrome” has emerged, emphasizing the putative role of sensitized sensory nerve endings and afferent neural pathways in causing chronic refractory cough, akin to chronic neuropathic pain. It presents with a dry or minimally productive cough and a tickle or sensitivity in the throat, made worse with talking, laughing, or exertion. It is more common in women than men and can last for years. Specific diagnostic criteria are lacking; the diagnosis is suspected when alternative etiologies are excluded by diagnostic testing or failed therapeutic trials. It is uncertain whether persistent daily coughing elicits an inflammatory response and is thereby self-perpetuating.

ACE inhibitor-induced cough occurs in 5–30% of patients taking these agents and is not dose-dependent. ACE metabolizes bradykinin and other tachykinins, such as substance P. The mechanism of ACE inhibitor-associated cough may involve sensitization of sensory nerve endings due to accumulation of bradykinin. Any patient with chronic unexplained cough who is taking an ACE inhibitor should have a trial period off the medication, regardless of the timing of the onset of cough relative to the initiation of ACE inhibitor therapy. In most instances, a safe alternative is available; angiotensin receptor blockers do not cause cough. Failure to observe a decrease in cough after 1 month off medication argues strongly against this etiology.

Postnasal drainage of any etiology can cause cough as a response to stimulation of sensory receptors of the cough-reflex pathway in the hypopharynx or aspiration of draining secretions into the trachea. The term *upper airway cough syndrome* has been coined to encompass the concept that chronic inflammation in the nose and sinuses can cause cough even in the absence of physical drainage into the pharynx. Historical clues suggesting this etiology include a sensation of postnasal drip, frequent throat clearing, and sneezing and rhinorrhea. On speculum examination of the nose, excess mucoi or purulent secretions, inflamed and edematous nasal mucosa, and/or polyps may be seen; in addition, secretions or a cobblestoned appearance of the mucosa along the posterior pharyngeal wall may be noted. Unfortunately, there is no means by which to quantitate postnasal drainage. In many instances, this diagnosis must rely on subjective information provided by the patient. Furthermore, this assessment must also be counterbalanced by the fact that many people who have chronic postnasal drainage do not experience cough.

Linking gastroesophageal reflux to chronic cough poses similar challenges. It is thought that reflux of gastric contents into the lower esophagus may trigger cough via reflex pathways initiated in the esophageal mucosa. Reflux to the level of the pharynx (laryngopharyngeal reflux), with consequent aspiration of gastric contents, causes a chemical bronchitis and possibly pneumonitis that can elicit cough for days afterward, but it is a rare finding among persons with chronic cough. Retrosternal burning after meals or on recumbency, frequent eructation, hoarseness, and throat pain may be indicative of gastroesophageal reflux. Nevertheless, reflux may also elicit minimal or no symptoms. Glottic inflammation detected on laryngoscopy may be a manifestation of recurrent reflux to the level of the throat, but it is a nonspecific finding. Quantification of the frequency and level of reflux requires a somewhat invasive procedure to measure esophageal pH (either nasopharyngeal placement of a catheter with a pH probe into the esophagus for 24 h or endoscopic placement of a radiotransmitter capsule into the esophagus) and, with newer techniques, esophageal pressures (manometry) and nonacid reflux. The precise interpretation of test results that permits an etiologic linking of reflux events and cough remains debated. Again, assigning the cause of cough to gastroesophageal reflux must be weighed against the observation that many people with symptomatic reflux do not experience chronic cough.

Cough alone as a manifestation of asthma is common among children but not among adults. Cough due to asthma in the absence of wheezing, shortness of breath, and chest tightness is referred to as “cough-variant asthma.” A history suggestive of cough-variant asthma

ties the onset of cough to exposure to typical triggers for asthma and the resolution of cough to discontinuation of exposure. Objective testing can establish the diagnosis of asthma (airflow obstruction on spirometry that varies over time or reverses in response to a bronchodilator) or exclude it with certainty (a negative response to a bronchoprovocation challenge—e.g., with methacholine). In a patient capable of taking reliable measurements, home expiratory peak flow monitoring can be a cost-effective method to support or discount a diagnosis of asthma.

Eosinophilic bronchitis causes chronic cough with a normal chest radiograph. This uncommon condition is characterized by sputum eosinophilia in excess of 3% without airflow obstruction or bronchial hyperresponsiveness and is successfully treated with inhaled glucocorticoids. Measurement of an elevated concentration of nitric oxide in exhaled breath has the potential to detect eosinophilic airway inflammation (in asthma or eosinophilic bronchitis) and predict a favorable response to inhaled steroids in persons with chronic cough.

Treatment of chronic cough in a patient with a normal chest radiograph is often empirical and is targeted at the most likely cause(s) of cough as determined by history, physical examination, and possibly pulmonary function testing. Therapy for postnasal drainage depends on the presumed etiology (infection, allergy, or vasomotor rhinitis) and may include systemic antihistamines; decongestants; antibiotics; nasal saline irrigation; and nasal pump sprays with glucocorticoids, antihistamines, or anticholinergics. Antacids histamine type 2 (H_2) receptor antagonists, and proton pump inhibitors are used to neutralize or decrease the production of gastric acid in gastroesophageal reflux disease; dietary changes, elevation of the head and torso during sleep, and medications to improve gastric emptying or impede the flow of refluxate (e.g., alginates) are additional therapeutic measures. Cough-variant asthma typically responds well to inhaled glucocorticoids and intermittent use of inhaled β -agonist bronchodilators.

Patients who fail to respond to treatment targeting the common causes of chronic cough or who have had these causes excluded by appropriate diagnostic testing should, in the opinion of the author, undergo chest CT. Diseases causing cough that may be missed on chest x-ray include tumors, early interstitial lung disease, bronchiectasis, and atypical mycobacterial pulmonary infection. On the other hand, patients with chronic cough who have normal findings on chest examination, lung function testing, oxygenation assessment, and chest CT can be reassured as to the absence of serious pulmonary pathology.

■ GLOBAL CONSIDERATIONS

Regular exposure to air pollution can cause chronic cough and throat clearing, as well as lower respiratory tract disease. Smoke from cooking and heating fuels in poorly ventilated homes; toxic exposures in work settings lacking implementation of occupational safety standards; and ambient chemicals and particulates in highly polluted outdoor air are all forms of air pollution causing cough. Limited therapeutic options are available; treatment focuses on improving environmental air quality (e.g., use of a stove chimney in the home), removal from the exposure, and use of an appropriate face mask.

In areas of the world where tuberculosis is endemic, chronic cough conjures the possibility of active pulmonary tuberculosis and mandates appropriate evaluation, including chest imaging and sputum analysis.

■ SYMPTOM-BASED TREATMENT OF COUGH

Empiric treatment of chronic idiopathic cough with inhaled corticosteroids, inhaled anticholinergic bronchodilators, and macrolide antibiotics has been tried without consistent success. Currently available cough suppressants are only modestly effective. Most potent are narcotic cough suppressants, such as codeine, hydrocodone, or morphine, which are thought to act in the “cough center” in the brainstem. The tendency of narcotic cough suppressants to cause drowsiness and constipation and their potential for addictive dependence limit their appeal for long-term use. Dextromethorphan is an over-the-counter, centrally acting cough suppressant with fewer side effects and less efficacy than the narcotic cough suppressants. Dextromethorphan is thought to have a different site of action than narcotic cough suppressants and can be used in combination with them if necessary. Benzonatate is thought to

inhibit neural activity of sensory nerves in the cough-reflex pathway. It is generally free of side effects; however, its effectiveness in suppressing cough is variable and unpredictable. Inhaled lidocaine, an inhibitor of voltage-gated sodium channels, provides transient cough suppression, but because of associated oropharyngeal anesthesia, it poses the risk of aspiration.

Attempts to treat cough hypersensitivity syndrome have focused on inhibition of neural pathways. Small case series and randomized clinical trials have indicated benefit from off-label use of gabapentin, pregabalin, or amitriptyline. Recent studies suggest a role for behavioral modification using specialized speech therapy techniques, but widespread application of this modality is currently not practical. Novel cough suppressants without the limitations of currently available agents are greatly needed. Approaches that are being explored include the development of neurokinin-1 receptor antagonists, transient receptor protein vanilloid-1 (TRPV1) channel antagonists, a promising P2X3 channel antagonist (gefapixant), and novel opioid and opioid-like receptor agonists.

FURTHER READING

- BRIGHTLING CE et al: Eosinophilic bronchitis as an important cause of chronic cough. *Am J Respir Crit Care Med* 160:406, 1999.
 CARROLL TL (ed): *Chronic Cough*. San Diego, Plural Publishing, Inc., 2019.
 GIBSON P et al: Treatment of unexplained chronic cough: CHEST guideline and expert panel report. *Chest* 149:27, 2016.
 KAHRILAS PJ et al: Chronic cough due to gastroesophageal reflux in adults: CHEST Guideline and Expert Panel Report. *Chest* 150:1381, 2016.
 MORICE AH et al: ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. *Eur Respir J* 55: 1901136, 2020.
 RAMSAY LE et al: Double-blind comparison of losartan, lisinopril and hydrochlorothiazide in hypertensive patients with previous angiotensin converting enzyme inhibitor-associated cough. *J Hypertens Suppl* 13:S73, 1995.
 RYAN NM et al: Gabapentin for refractory chronic cough: A randomized, double-blind, placebo-controlled trial. *Lancet* 380:1583, 2012.
 SMITH JA, WOODCOCK A: Chronic cough. *N Engl J Med* 375:1544, 2016.

39

Hemoptysis

Carolyn M. D'Ambrosio



Hemoptysis is the expectoration of blood from the respiratory tract. Bleeding from the gastrointestinal tract (hematemesis) or nasal cavities (epistaxis) can mimic hemoptysis. Once established as hemoptysis, the degree of blood that is being expectorated (volume and frequency) is the next step as massive or life-threatening hemoptysis (>400 mL of blood in 24 h or >150 mL at one time) requires emergent intervention. This chapter will focus predominantly on non-life-threatening hemoptysis. The source of the bleeding as well as the cause are the next steps when approaching a patient with hemoptysis.

ANATOMY AND PHYSIOLOGY OF HEMOPTYSIS

Hemoptysis can arise from anywhere in the respiratory tract, from the glottis to the alveolus. Most commonly, bleeding arises from the bronchi or medium-sized airways, but a thorough evaluation of the entire respiratory tree is important.

The dual blood supply of the lungs makes it unique. The lungs have both the pulmonary and bronchial circulations. The pulmonary circulation is a low-pressure system that is essential for gas exchange at the alveolar level; in contrast, the bronchial circulation originates from the aorta and, therefore, is a higher-pressure system. The bronchial arteries supply the airways and can neovascularize tumors, dilated airways of bronchiectasis, and cavitary lesions. Most hemoptysis originates from the bronchial circulation, and bleeding from the higher-pressure system makes it more difficult to stop.

ETIOLOGY

Hemoptysis commonly results from infection, malignancy, or vascular disease; however, the differential for bleeding from the respiratory tree is varied and broad. In the United States, the most common causes are viral bronchitis, bronchiectasis, or malignancy. In other parts of the world, infections such as tuberculosis are the most common causes.

Infections Most blood-tinged sputum and small-volume hemoptysis are due to viral bronchitis. Patients with chronic bronchitis are at risk for bacterial superinfection with organisms such as *Streptococcus pneumoniae*, *Haemophilus influenzae*, or *Moraxella catarrhalis*, increasing airway inflammation and potential for bleeding. Similarly, patients with bronchiectasis are prone to hemoptysis during exacerbations. Due to recurrent bacterial infection, bronchiectatic airways are dilated, inflamed, and highly vascular, supplied by the bronchial circulation. In several case series, bronchiectasis is the leading cause of massive hemoptysis and subsequent death.

Tuberculosis had long been the most common cause of hemoptysis worldwide, but it is now surpassed in industrialized countries by bronchitis and bronchiectasis. In patients with tuberculosis, development of cavitary disease is frequently the source of bleeding, but rarer complications such as the erosion of a pulmonary artery aneurysm into a preexisting cavity (i.e., Rasmussen's aneurysm) can also be the source.

Other infectious agents such as endemic fungi, *Nocardia*, and non-tuberculous mycobacteria can present as cavitary lung disease complicated by hemoptysis. In addition, *Aspergillus* species can develop into mycetomas within preexisting cavities, with neovascularization to these inflamed spaces leading to bleeding. Pulmonary abscesses and necrotizing pneumonia can cause bleeding by devitalizing lung parenchyma. Common responsible organisms include *Staphylococcus aureus*, *Klebsiella pneumoniae*, and oral anaerobes.

Paragonimiasis can mimic tuberculosis and is another significant cause of hemoptysis seen globally; it is common in Southeast Asia and China, although cases have been reported in North America from raw crayfish ingestion. It should be considered as a cause of hemoptysis in recent immigrants from endemic areas.

Vascular Hemoptysis from a vascular cause can be associated with cardiac disease, pulmonary embolism, arteriovenous malformation, or diffuse alveolar hemorrhage (DAH). While the classic description of the sputum expectorated in pulmonary edema (from elevated left end-diastolic pressure) is "pink and frothy," a spectrum of hemoptysis including frank blood can be seen. This observation is particularly true now with the more widespread use of anticoagulants and antiplatelet medications.

Pulmonary embolism with parenchymal infarction can present with hemoptysis, but pulmonary emboli do not commonly cause hemoptysis. An ectatic vessel in an airway or a pulmonary arteriovenous malformation can be a source of bleeding. A rare vascular cause of hemoptysis is the rupture of an aortobronchial fistula; these fistulae arise in the setting of aortic pathology such as aneurysm or pseudoaneurysm and can cause small bleeding episodes that herald massive hemoptysis.

DAH causes significant bleeding into the lung parenchyma but, interestingly, is not often associated with hemoptysis. DAH typically presents with diffuse ground glass opacities on chest imaging. A range of insults cause DAH, including immune-mediated capillaritis from diseases such as systemic lupus erythematosus, toxicity from cocaine and other inhalants, and stem cell transplantation. The

so-called “pulmonary-renal” syndromes, including granulomatosis with polyangiitis and anti-glomerular basement membrane disease, may lead to both hemoptysis and hematuria (though one manifestation may be present without the other). A recently identified cause of hemoptysis and DAH is vaping-induced lung injury.

Malignancy Bronchogenic carcinoma of any histology is a common cause of hemoptysis (both massive and nonmassive). Hemoptysis can indicate airway involvement of the tumor and can be a presenting symptom of carcinoid tumors, vascular lesions that frequently arise in the proximal airways. Small cell and squamous cell carcinomas are frequently central in nature and more likely to erode into major pulmonary vessels, resulting in massive hemoptysis. Pulmonary metastases from distant tumors (e.g., melanoma, sarcoma, adenocarcinomas of the breast and colon) can also cause bleeding. Kaposi’s sarcoma, seen in advanced acquired immunodeficiency syndrome, is very vascular and can develop anywhere along the respiratory tract, from the bronchi to the oral cavity.

Mechanical and Other Causes In addition to infection, vascular disease, and malignancy, other insults to the pulmonary system can cause hemoptysis. Pulmonary endometriosis causes cyclical bleeding known as catamenial hemoptysis. Foreign body aspiration can lead to airway irritation and bleeding. Diagnostic and therapeutic procedures are also potential offenders: pulmonary vein stenosis can result from left atrial procedures, such as pulmonary vein isolation, and pulmonary artery catheters can lead to rupture of the pulmonary artery if the distal balloon is kept inflated. Finally, in the setting of thrombocytopenia, coagulopathy, anticoagulation, or antiplatelet therapy, even minor insults can cause hemoptysis.

EVALUATION AND MANAGEMENT

History The amount or severity of bleeding is the first step in assessing a patient with hemoptysis. A patient’s description of the sputum (e.g., flecks of blood, pink-tinged, or frank blood or clot) is helpful if you cannot examine it. An approach to management of hemoptysis is outlined in Fig. 39-1.

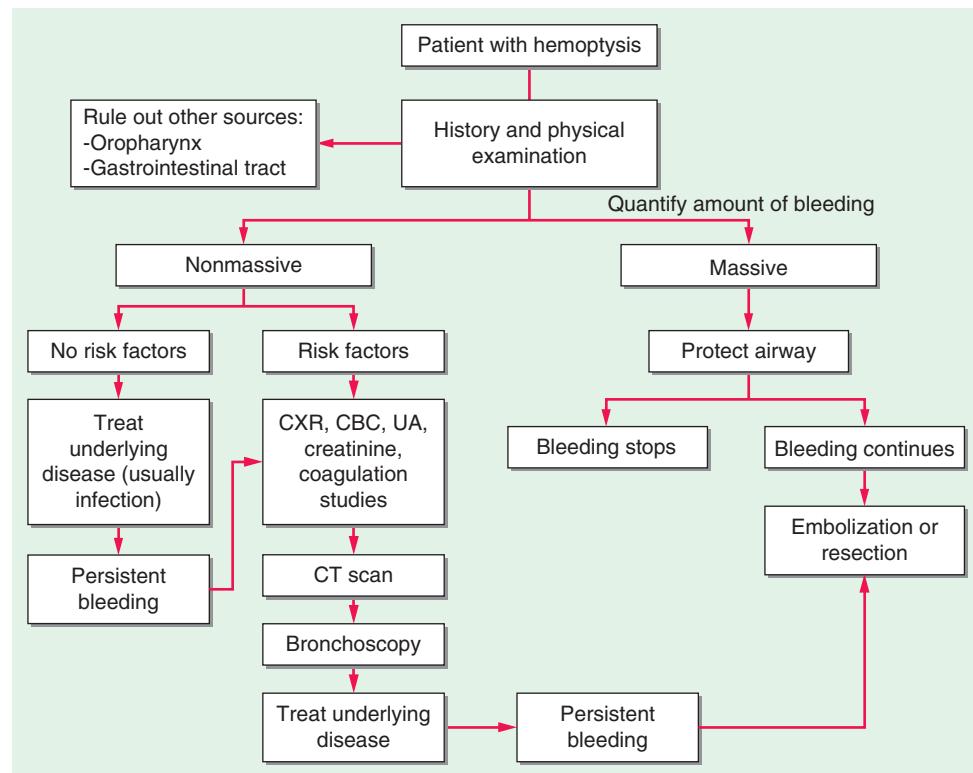


FIGURE 39-1 Approach to the management of hemoptysis. CBC, complete blood count; CT, computed tomography; CXR, chest x-ray; UA, urinalysis.

While there is no agreed-upon volume, blood loss of 400 mL in 24 h or 100–150 mL expectorated at one time should be considered *life-threatening hemoptysis*. These numbers derive from the blood volume of the tracheobronchial tree (generally 100–200 mL). Patients rarely die of exsanguination but, rather, are at risk of death due to asphyxiation from blood filling the airways and airspaces. Most patients cannot describe the volume of their hemoptysis in milliliters, so using referents like cups (one U.S. cup is 236 mL) can be helpful. Fortunately, life-threatening hemoptysis only accounts for 5–15% of cases of hemoptysis.

The history may point to the cause of hemoptysis. Fever, chills, or antecedent cough may suggest infection. A history of smoking or unintentional weight loss makes malignancy more likely. Patients should be asked about inhalational exposures, including vaping. A thorough medical history with careful attention to chronic pulmonary disease should be obtained, with evaluation of risk factors for malignancy and bronchiectatic lung disease (e.g., cystic fibrosis, sarcoidosis).

Physical Examination Reviewing the vital signs is an important first step. Patients who have life-threatening hemoptysis can have hypoxemia, tachycardia, and hemodynamic instability. As the site of bleeding is important, evaluation of the nasal and oral cavities is imperative. In addition, auscultation of the lungs and seeking other relevant physical findings such as clubbing can point to a cause of the hemoptysis. A focal area of wheezing could suggest a foreign body aspiration. Other signs of a bleeding diathesis (e.g., skin or mucosal ecchymoses and petechiae) or telangiectasias may suggest other etiologies of the hemoptysis.

Diagnostic Studies Initial studies should include measurement of a complete blood count to assess for infection, anemia, or thrombocytopenia; coagulation parameters; measurement of electrolytes and renal function; and urinalysis to exclude pulmonary-renal disease. Chest imaging is necessary for every patient.

A chest radiograph is usually obtained first, although it frequently does not localize bleeding and can appear normal. In patients without risk factors for malignancy or other abnormalities in the initial evaluation and with a normal chest radiograph, treating for bronchitis and ensuring close follow-up is a reasonable strategy, with further diagnostic workup.

In contrast, patients with risk factors for malignancy (i.e., age >40 or a smoking history) should undergo additional testing. First, chest computed tomography (CT) with contrast should be obtained to better identify masses, bronchiectasis, and parenchymal lesions. A CT looking for pulmonary embolism should be considered if the history and physical examination are consistent with that diagnosis. Following a CT, a flexible bronchoscopy should be performed to exclude bronchogenic carcinoma unless imaging reveals a lesion that can be sampled without bronchoscopy. Small case series show that patients with hemoptysis and unrevealing bronoscopies have good outcomes.

Interventions When the amount of hemoptysis is massive or life-threatening, there are three simultaneous goals: first, protect the nonbleeding lung; second, locate the site of bleeding; and third, control the bleeding.

Protecting the airway and nonbleeding lung is paramount in the management of massive hemoptysis because asphyxiation can happen quickly. If the side of bleeding is known, the patient should be positioned with the bleeding side down to use gravitational advantage to keep blood out of the nonbleeding lung. Endotracheal intubation should be avoided unless truly necessary, since suctioning through an endotracheal tube is a less effective means of removing blood and clot than the cough reflex. If intubation is required, take steps to protect the nonbleeding lung either by selective intubation of one lung (i.e., the nonbleeding lung) or insertion of a double-lumen endotracheal tube.

Locating the bleeding site is sometimes obvious, but frequently, it can be difficult to determine. A chest radiograph, if it shows new opacities, can be helpful in localizing the side or site of bleeding, although this test is not adequate by itself. CT angiography helps by localizing active extravasation. Flexible bronchoscopy may be useful to identify the side of bleeding (although it has only a 50% chance of locating the site). Experts do not agree on the timing of bronchoscopy, although in some cases—cystic fibrosis, for instance—bronchoscopy is *not* recommended because it may delay definitive management. Finally, proceeding directly to angiography is also a reasonable strategy given that it has both diagnostic and therapeutic capabilities.

Controlling the bleeding during an episode of life-threatening hemoptysis can be accomplished in one of three ways: from the airway lumen, from the involved blood vessel, or by surgical resection of both airway and vessel involved. Bronchoscopic measures are generally only temporizing: a flexible bronchoscope can be used to suction clot and insert a balloon catheter or bronchial blocker that occludes the involved airway. Rigid bronchoscopy, done by an interventional pulmonologist or thoracic surgeon, may allow therapeutic interventions of bleeding airway lesions such as photocoagulation and cauterization. Because most life-threatening cases of hemoptysis arise from the bronchial circulation, bronchial artery embolization is the procedure of choice for control of the bleeding. However, bronchial artery embolization can have significant complications such as embolization of the anterior spinal artery. However, it is generally successful in the short term, with >80% success rate at controlling bleeding immediately, although bleeding can recur if the underlying disease (e.g., a mycetoma) is not treated. Surgical resection has a high mortality rate (up to 15–40%) and should not be pursued unless initial measures have failed and bleeding is ongoing. Ideal candidates for surgery have localized disease but otherwise normal lung parenchyma.

ACKNOWLEDGMENT

Anna K. Brady and Patricia A. Kritek contributed to this chapter in the 20th edition, and some material from that chapter has been retained here.

FURTHER READING

- ADELMAN M et al: Cryptogenic hemoptysis: Clinical features, bronchoscopic findings, and natural history in 67 patients. Ann Intern Med 102:829, 1985.
- FLUME PA et al: CF pulmonary guidelines. Pulmonary complications: Hemoptysis and pneumothorax. Am J Respir Crit Care Med 182:298, 2010.
- HIRSHBERG B et al: Hemoptysis: Etiology, evaluation, and outcome in a tertiary care hospital. Chest 112:440, 1997.
- JEAN-BAPTISTE E: Clinical assessment and management of massive hemoptysis. Crit Care Med 28:1642, 2000.
- JOHNSON JL: Manifestations of hemoptysis: How to manage minor, moderate, and massive bleeding. Postgrad Med 112:4:101, 2002.
- LAYDEN JE et al: Pulmonary illness related to e-cigarette, reply. N Engl J Med 382:903, 2020.
- LORDAN JL et al: The pulmonary physician in critical care: Illustrative case 7. Assessment and management of massive hemoptysis. Thorax 58:814, 2003.
- SOPKO DR, SMITH TP: Bronchial artery embolization for massive hemoptysis. Semin Intervent Radiol 28:48, 2011.

40

Hypoxia and Cyanosis

Joseph Loscalzo



HYPOXIA

The fundamental purpose of the cardiorespiratory system is to deliver O₂ and nutrients to cells and to remove CO₂ and other metabolic products from them. Proper maintenance of this function depends not only on intact cardiovascular and respiratory systems, but also on an adequate number of red blood cells and hemoglobin and a supply of inspired gas containing adequate O₂.

■ RESPONSES TO HYPOXIA

Decreased O₂ availability to cells typically results in an inhibition of oxidative phosphorylation and increased anaerobic glycolysis. This switch from aerobic to anaerobic metabolism, the Pasteur effect, reduces the rate of adenosine 5'-triphosphate (ATP) production. In severe hypoxia, when ATP production is inadequate to meet the energy requirements of ionic and osmotic equilibrium, cell membrane depolarization leads to uncontrolled Ca²⁺ influx and activation of Ca²⁺-dependent phospholipases and proteases. These events, in turn, cause cell swelling, activation of apoptotic pathways, and, ultimately, cell death.

The adaptations to hypoxia are mediated, in part, by the upregulation of genes encoding a variety of proteins, including glycolytic enzymes, such as phosphoglycerate kinase and phosphofructokinase, as well as the glucose transporters Glut-1 and Glut-2; and by growth factors, such as vascular endothelial growth factor (VEGF) and erythropoietin, which enhance erythrocyte production. The hypoxia-induced increase in expression of these and other key proteins is governed by the hypoxia-sensitive transcription factor, hypoxia-inducible factor-1 (HIF-1).

During hypoxia, systemic arterioles dilate, at least in part, by opening of K_{ATP} channels in vascular smooth-muscle cells due to the hypoxia-induced reduction in ATP concentration. By contrast, in pulmonary vascular smooth-muscle cells, inhibition of K⁺ channels causes depolarization, which, in turn, activates voltage-gated Ca²⁺ channels, raising the cytosolic [Ca²⁺] and causing smooth-muscle cell contraction. Hypoxia-induced pulmonary arterial constriction shunts blood away from poorly ventilated portions toward better ventilated portions of the lung (i.e., improves ventilation-perfusion mismatch); however, it also increases pulmonary vascular resistance and right ventricular afterload.

Effects on the Central Nervous System Changes in the central nervous system (CNS), particularly the higher centers, are especially important consequences of hypoxia. Acute hypoxia causes impaired judgment, motor incoordination, and a clinical picture resembling acute alcohol intoxication. High-altitude illness is characterized by headache secondary to cerebral vasodilation, gastrointestinal symptoms, dizziness, insomnia, fatigue, or somnolence. Pulmonary arterial and sometimes venous constriction causes capillary leakage and high-altitude pulmonary edema (HAPE) ([Chap. 37](#)), which intensifies hypoxia, further promoting vasoconstriction. Rarely, high-altitude cerebral edema (HACE) develops, which is manifest by severe headache and papilledema and can cause coma. As hypoxia becomes more severe, the regulatory centers of the brainstem are affected, and death usually results from respiratory failure.

Effects on the Cardiovascular System Acute hypoxia stimulates the chemoreceptor reflex arc to induce vasoconstriction and systemic arterial vasodilation. These acute changes are accompanied by transiently increased myocardial contractility, which is followed by depressed myocardial contractility with prolonged hypoxia.

■ CAUSES OF HYPOXIA

Respiratory Hypoxia When hypoxia occurs from respiratory failure, Pao₂ declines, and when respiratory failure is persistent, the

hemoglobin-oxygen (Hb-O_2) dissociation curve (see Fig. 98-2) is displaced to the right, with greater quantities of O_2 released at any level of tissue Po_2 . Arterial hypoxemia, that is, a reduction of O_2 saturation of arterial blood (SaO_2), and consequent cyanosis are likely to be more marked when such depression of Pao_2 results from pulmonary disease than when the depression occurs as the result of a decline in the fraction of oxygen in inspired air (FiO_2). In this latter situation, Paco_2 falls secondary to anoxia-induced hyperventilation and the Hb-O_2 dissociation curve is displaced to the left, limiting the decline in SaO_2 at any level of Pao_2 .

The most common cause of respiratory hypoxia is *ventilation-perfusion mismatch* resulting from perfusion of poorly ventilated alveoli. Respiratory hypoxemia may also be caused by *hypoventilation*, in which case it is associated with an elevation of Paco_2 (Chap. 285). These two forms of respiratory hypoxia are usually correctable by inspiring 100% O_2 for several minutes. A third cause of respiratory hypoxia is shunting of blood across the lung from the pulmonary arterial to the venous bed (*intrapulmonary right-to-left shunting*) by perfusion of nonventilated portions of the lung, as in pulmonary atelectasis or through pulmonary arteriovenous connections. The low Pao_2 in this situation is only partially corrected by an FiO_2 of 100%.

Hypoxia Secondary to High Altitude As one ascends rapidly to 3000 m (~10,000 ft), the reduction of the O_2 content of inspired air (FiO_2) leads to a decrease in alveolar Po_2 to ~60 mmHg, and a condition termed *high-altitude illness* develops (see above). At higher altitudes, arterial saturation declines rapidly and symptoms become more serious; and at 5000 m, unacclimated individuals usually cease to be able to function normally owing to the changes in CNS function described above.

Hypoxia Secondary to Right-to-Left Extrapulmonary Shunting From a physiologic viewpoint, this cause of hypoxia resembles intrapulmonary right-to-left shunting but is caused by congenital cardiac malformations, such as tetralogy of Fallot, transposition of the great arteries, atrial or ventricular septal defect, patent ductus arteriosus, and Eisenmenger's syndrome (Chap. 269). As in pulmonary right-to-left shunting, the Pao_2 cannot be restored to normal with inspiration of 100% O_2 .

Anemic Hypoxia A reduction in hemoglobin concentration of the blood is accompanied by a corresponding decline in the O_2 -carrying capacity of the blood. Although the Pao_2 is normal in anemic hypoxia, the absolute quantity of O_2 transported per unit volume of blood is diminished. As the anemic blood passes through the capillaries and the usual quantity of O_2 is removed from it, the Po_2 and saturation in the venous blood decline to a greater extent than normal.

Carbon Monoxide (CO) Intoxication (See also Chap. 463) Hemoglobin that binds with CO (carboxy-hemoglobin [COHb]) is unavailable for O_2 transport. In addition, the presence of COHb shifts the Hb-O_2 dissociation curve to the left (see Fig. 98-2) so that O_2 is unloaded only at lower tensions, further contributing to tissue hypoxia.

Circulatory Hypoxia As in anemic hypoxia, the Pao_2 is usually normal, but venous and tissue Po_2 values are reduced as a consequence of reduced tissue perfusion and greater tissue O_2 extraction. This pathophysiology leads to an increased arterial-mixed venous O_2 difference ($a\text{-v-O}_2$ difference), or gradient. Generalized circulatory hypoxia occurs in heart failure (Chap. 257) and in most forms of shock (Chap. 303).

Specific Organ Hypoxia Localized circulatory hypoxia may occur as a result of decreased perfusion secondary to arterial obstruction, as in localized atherosclerosis in any vascular bed, or as a consequence of vasoconstriction, as observed in Raynaud's phenomenon (Chap. 281). Localized hypoxia may also result from venous obstruction and the resultant expansion of interstitial fluid causing arteriolar compression and, thereby, reduction of arterial inflow. Edema, which increases the distance through which O_2 must diffuse before it reaches cells, can also cause localized hypoxia. In an attempt to maintain adequate perfusion to more vital organs in patients with reduced cardiac output secondary

to heart failure or hypovolemic shock, vasoconstriction may reduce perfusion in the limbs and skin, causing hypoxia of these regions.

Increased O_2 Requirements If the O_2 consumption of tissues is elevated without a corresponding increase in perfusion, tissue hypoxia ensues and the Po_2 in venous blood declines. Ordinarily, the clinical picture of patients with hypoxia due to an elevated metabolic rate, as in fever or thyrotoxicosis, is quite different from that in other types of hypoxia: the skin is warm and flushed owing to increased cutaneous blood flow that dissipates the excessive heat produced, and cyanosis is usually absent.

Exercise is a classic example of increased tissue O_2 requirements. These increased demands are normally met by several mechanisms operating simultaneously: (1) increase in the cardiac output and ventilation and, thus, O_2 delivery to the tissues; (2) a preferential shift in blood flow to the exercising muscles by changing vascular resistances in the circulatory beds of exercising tissues, directly and/or reflexly; (3) an increase in O_2 extraction from the delivered blood and a widening of the arteriovenous O_2 difference; and (4) a reduction in the pH of the tissues and capillary blood, shifting the Hb-O_2 curve to the right (see Fig. 98-2), and unloading more O_2 from hemoglobin. If the capacity of these mechanisms is exceeded, then hypoxia, especially of the exercising muscles, will result.

Improper Oxygen Utilization Cyanide (Chap. 459) and several other similarly acting poisons cause cellular hypoxia by impairing electron transport in mitochondria, thereby limiting oxidative phosphorylation and ATP production. The tissues are unable to use O_2 , and as a consequence, the venous blood tends to have a high O_2 tension. This condition has been termed *histotoxic hypoxia*.

■ ADAPTATION TO HYPOXIA

An important component of the respiratory response to hypoxia originates in special chemosensitive cells in the carotid and aortic bodies and in the respiratory center in the brainstem. The stimulation of these cells by hypoxia increases ventilation, with a loss of CO_2 , and can lead to respiratory alkalosis. When combined with the metabolic acidosis resulting from the production of lactic acid, the serum bicarbonate level declines (Chap. 55).

With the reduction of Pao_2 , cerebrovascular resistance decreases and cerebral blood flow increases in an attempt to maintain O_2 delivery to the brain. However, when the reduction of Pao_2 is accompanied by hyperventilation and a reduction of Paco_2 , cerebrovascular resistance rises, cerebral blood flow falls, and tissue hypoxia intensifies.

The diffuse, systemic vasodilation that occurs in generalized hypoxia increases the cardiac output. In patients with underlying heart disease, the requirements of peripheral tissues for an increase of cardiac output with hypoxia may precipitate congestive heart failure. In patients with ischemic heart disease, a reduced Pao_2 may intensify myocardial ischemia and further impair left ventricular function.

One of the important compensatory mechanisms for chronic hypoxia is an increase in the hemoglobin concentration and in the number of red blood cells in the circulating blood, that is, the development of polycythemia induced by erythropoietin production (Chap. 103). In persons with chronic hypoxemia secondary to prolonged residence at a high altitude (>13,000 ft, 4200 m), a condition termed *chronic mountain sickness* develops. This disorder is characterized by a blunted respiratory drive, reduced ventilation, erythrocytosis, cyanosis, weakness, right ventricular enlargement secondary to pulmonary hypertension, and even stupor.

CYANOSIS

Cyanosis refers to a bluish color of the skin and mucous membranes resulting from an increased quantity of reduced hemoglobin (i.e., deoxygenated hemoglobin) or of hemoglobin derivatives (e.g., methemoglobin or sulfhemoglobin) in the small blood vessels of those tissues. It is usually most marked in the lips, nail beds, ears, and malar eminences. Cyanosis, especially if developed recently, is more commonly detected by a family member than the patient. The florid skin characteristic of polycythemia vera (Chap. 103) must be distinguished from the true cyanosis discussed here. A cherry-colored flush, rather than cyanosis, is caused by COHb (Chap. 459).

The degree of cyanosis is modified by the color of the cutaneous pigment and the thickness of the skin, as well as by the state of the cutaneous capillaries. The accurate clinical detection of the presence and degree of cyanosis is difficult, as proved by oximetric studies. In some instances, central cyanosis can be detected reliably when the Sao_2 has fallen to 85%; in others, particularly in dark-skinned persons, it may not be detected until it has declined to 75%. In the latter case, examination of the mucous membranes in the oral cavity and the conjunctivae rather than examination of the skin is more helpful in the detection of cyanosis.

The increase in the quantity of reduced hemoglobin in the mucocutaneous vessels that produces cyanosis may be brought about either by an increase in the quantity of venous blood as a result of dilation of the venules (including precapillary venules) or by a reduction in the Sao_2 in the capillary blood. In general, cyanosis becomes apparent when the concentration of reduced hemoglobin in capillary blood exceeds 40 g/L (4 g/dL).

It is the *absolute*, rather than the *relative*, quantity of reduced hemoglobin that is important in producing cyanosis. Thus, in a patient with severe anemia, the *relative* quantity of reduced hemoglobin in the venous blood may be very large when considered in relation to the total quantity of hemoglobin in the blood. However, since the concentration of the latter is markedly reduced, the *absolute* quantity of reduced hemoglobin may still be low, and, therefore, patients with severe anemia and even *marked* arterial desaturation may not display cyanosis. Conversely, the higher the total hemoglobin content, the greater is the tendency toward cyanosis; thus, patients with marked polycythemia tend to be cyanotic at higher levels of Sao_2 than patients with normal hematocrit values. Likewise, local passive congestion, which causes an increase in the total quantity of reduced hemoglobin in the vessels in a given area, may cause cyanosis. Cyanosis is also observed when nonfunctional hemoglobin, such as methemoglobin (consequential or acquired) or sulfhemoglobin (**Chap. 98**), is present in blood.

Cyanosis may be subdivided into central and peripheral types. In *central cyanosis*, the Sao_2 is reduced or an abnormal hemoglobin derivative is present, and the mucous membranes and skin are both affected. *Peripheral cyanosis* is due to a slowing of blood flow and abnormally great extraction of O_2 from normally saturated arterial blood; it results from vasoconstriction and diminished peripheral blood flow, such as occurs in cold exposure, shock, congestive failure, and peripheral vascular disease. Often in these conditions, the mucous membranes of the oral cavity, including the sublingual mucosa, may be spared. Clinical differentiation between central and peripheral cyanosis may not always be straightforward, and in conditions such as cardiogenic shock with pulmonary edema, there may be a mixture of both types.

■ DIFFERENTIAL DIAGNOSIS

Central Cyanosis (Table 40-1) Decreased Sao_2 results from a marked reduction in the Pao_2 . This reduction may be brought about by a decline in the FiO_2 without sufficient compensatory alveolar hyperventilation to maintain alveolar Po_2 . Cyanosis usually becomes manifest in an ascent to an altitude of 4000 m (13,000 ft).

Seriously *impaired pulmonary function*, through perfusion of unventilated or poorly ventilated areas of the lung or alveolar hypoventilation, is a common cause of central cyanosis (**Chap. 285**). This condition may occur acutely, as in extensive pneumonia or pulmonary edema, or chronically, with chronic pulmonary diseases (e.g., emphysema). In the latter situation, secondary polycythemia is generally present and clubbing of the fingers (see below) may occur. Another cause of reduced Sao_2 is *shunting of systemic venous blood into the arterial circuit*. Certain forms of congenital heart disease are associated with cyanosis on this basis (see above and **Chap. 269**).

Pulmonary arteriovenous fistulae may be congenital or acquired, solitary or multiple, and microscopic or massive. The severity of cyanosis produced by these fistulae depends on their size and number. They occur with some frequency in hereditary hemorrhagic telangiectasia. Sao_2 reduction and cyanosis may also occur in some patients with cirrhosis, presumably as a consequence of pulmonary arteriovenous fistulae or portal vein-pulmonary vein anastomoses.

TABLE 40-1 Causes of Cyanosis

Central Cyanosis

- Decreased arterial oxygen saturation
- Decreased atmospheric pressure—high altitude
- Impaired pulmonary function
 - Alveolar hypoventilation
 - Inhomogeneity in pulmonary ventilation and perfusion (perfusion of hypoventilated alveoli)
 - Impaired oxygen diffusion
- Anatomic shunts
- Certain types of congenital heart disease
- Pulmonary arteriovenous fistulas
- Multiple small intrapulmonary shunts
- Hemoglobin with low affinity for oxygen
- Hemoglobin abnormalities
 - Methemoglobinemia—hereditary, acquired
 - Sulfhemoglobinemia—acquired
 - Carboxyhemoglobinemia (not true cyanosis)

Peripheral Cyanosis

- Reduced cardiac output
- Cold exposure
- Redistribution of blood flow from extremities
- Arterial obstruction
- Venous obstruction

In patients with cardiac or pulmonary right-to-left shunts, the presence and severity of cyanosis depend on the size of the shunt relative to the systemic flow and on the Hb-O_2 saturation of the venous blood. With increased extraction of O_2 from the blood by the exercising muscles, the venous blood returning to the right side of the heart is more unsaturated than at rest, and shunting of this blood intensifies the cyanosis. Secondary polycythemia occurs frequently in patients in this setting and contributes to the cyanosis.

Cyanosis can be caused by small quantities of circulating methemoglobin (Hb Fe^{3+}) and by even smaller quantities of sulfhemoglobin (**Chap. 98**); both of these hemoglobin derivatives impair oxygen delivery to the tissues. Although they are uncommon causes of cyanosis, these abnormal hemoglobin species should be sought by spectroscopy when cyanosis is not readily explained by malfunction of the circulatory or respiratory systems. Generally, digital clubbing does not occur with them.

Peripheral Cyanosis Probably the most common cause of peripheral cyanosis is the normal vasoconstriction resulting from exposure to cold air or water. When cardiac output is reduced, cutaneous vasoconstriction occurs as a compensatory mechanism so that blood is diverted from the skin to more vital areas such as the CNS and heart, and cyanosis of the extremities may result even though the arterial blood is normally saturated.

Arterial obstruction to an extremity, as with an embolus, or arteriolar constriction, as in cold-induced vasospasm (Raynaud's phenomenon) (**Chap. 281**), generally results in pallor and coldness, and there may be associated cyanosis. Venous obstruction, as in thrombophlebitis or deep venous thrombosis, dilates the subpapillary venous plexuses and thereby intensifies cyanosis.

APPROACH TO THE PATIENT

Cyanosis

Certain features are important in arriving at the cause of cyanosis:

1. It is important to ascertain the time of onset of cyanosis. Cyanosis present since birth or infancy is usually due to congenital heart disease.

2. Central and peripheral cyanosis must be differentiated. Evidence of disorders of the respiratory or cardiovascular systems is helpful. Massage or gentle warming of a cyanotic extremity will increase peripheral blood flow and abolish peripheral, but not central, cyanosis.
3. The presence or absence of clubbing of the digits (see below) should be ascertained. The combination of cyanosis and clubbing is frequent in patients with congenital heart disease and right-to-left shunting and is seen occasionally in patients with pulmonary disease, such as lung abscess or pulmonary arteriovenous fistulae. In contrast, peripheral cyanosis or acutely developing central cyanosis is *not* associated with clubbed digits.
4. Pao₂ and SaO₂ should be determined, and in patients with cyanosis in whom the mechanism is obscure, spectroscopic examination of the blood should be performed to look for abnormal types of hemoglobin (critical in the differential diagnosis of cyanosis).

CLUBBING

The selective bulbous enlargement of the distal segments of the fingers and toes due to proliferation of connective tissue, particularly on the dorsal surface, is termed *clubbing*; there is also increased sponginess of the soft tissue at the base of the clubbed nail. Clubbing may be hereditary, idiopathic, or acquired and associated with a variety of disorders, including cyanotic congenital heart disease (see above), infective endocarditis, and a variety of pulmonary conditions (among them primary and metastatic lung cancer, bronchiectasis, asbestos, sarcoidosis, lung abscess, cystic fibrosis, tuberculosis, and mesothelioma), as well as with some gastrointestinal diseases (including inflammatory bowel disease and hepatic cirrhosis). In some instances, it is occupational, for example, in jackhammer operators.

Clubbing in patients with primary and metastatic lung cancer, mesothelioma, bronchiectasis, or hepatic cirrhosis may be associated with *hypertrophic osteoarthropathy*. In this condition, the subperiosteal formation of new bone in the distal diaphyses of the long bones of the extremities causes pain and symmetric arthritis-like changes in the shoulders, knees, ankles, wrists, and elbows. The diagnosis of hypertrophic osteoarthropathy may be confirmed by bone radiograph or magnetic resonance imaging (MRI). Although the mechanism of clubbing is unclear, it appears to be secondary to humoral substances that cause dilation of the vessels of the distal digits as well as growth factors released from platelet precursors in the digital circulation. In certain circumstances, clubbing is reversible, such as following lung transplantation for cystic fibrosis.

FURTHER READING

- CALLEMEYN J et al: Clubbing and hypertrophic osteoarthropathy: Insights into diagnosis, pathophysiology, and clinical significance. *Acta Clin Belg* 22:1, 2016.
- MACINTYRE NR: Tissue hypoxia: Implications for the respiratory clinician. *Respir Care* 59:1590, 2014.

There is constant interchange of fluid between the two compartments of the extracellular fluid. The hydrostatic pressure within the capillaries and the colloid oncotic pressure in the interstitial fluid promote the movement of water and diffusible solutes from plasma to the interstitium. This movement is most prominent at the arterial origin of the capillary and falls progressively with the decline in intracapillary pressure and the rise in oncotic pressure toward the venular end. Fluid is returned from the interstitial space into the vascular system largely through the lymphatic system. These interchanges of fluids are normally balanced so that the volumes of the intravascular and interstitial compartments remain constant. However, a net movement of fluid from the intravascular to the interstitial spaces takes place and may be responsible for the development of edema under the following conditions: (1) an increase in intracapillary hydrostatic pressure; (2) inadequate lymphatic drainage; (3) reductions in the oncotic pressure in the plasma; (4) damage to the capillary endothelial barrier; and (5) increases in the oncotic pressure in the interstitial space.

REDUCTION OF EFFECTIVE ARTERIAL VOLUME

In many forms of edema, the effective arterial blood volume, a parameter that represents the filling of the arterial tree and that effectively perfuses the tissues, is reduced. Underfilling of the arterial tree may be caused by a reduction of cardiac output and/or systemic vascular resistance, by the pooling of blood in the splanchnic veins (as in cirrhosis), and by hypoalbuminemia (Fig. 41-1A). As a consequence of this underfilling, a series of physiologic responses designed to restore the effective arterial volume to normal are set into motion. A key element of these responses is the renal retention of sodium and, therefore, water, thereby restoring effective arterial volume, but sometimes also leading to the development or intensification of edema.

RENAL FACTORS AND THE RENIN-ANGIOTENSIN-ALDOSTERONE SYSTEM

The diminished renal blood flow characteristic of states in which the effective arterial blood volume is reduced is translated by the renal juxtaglomerular cells (specialized myoepithelial cells surrounding the afferent arteriole) into a signal for increased renin release. Renin is an enzyme with a molecular mass of about 40,000 Da that acts on its substrate, angiotensinogen, an α_2 -globulin synthesized by the liver, to release angiotensin I, a decapeptide, which in turn is converted to angiotensin II (AII), an octapeptide. AII has generalized vasoconstrictor properties, particularly on the renal efferent arterioles. This action reduces the hydrostatic pressure in the peritubular capillaries, whereas the increased filtration fraction raises the colloid osmotic pressure in these vessels, thereby enhancing salt and water reabsorption in the proximal tubule as well as in the ascending limb of the loop of Henle.

The renin-angiotensin-aldosterone system (RAAS) operates as both a hormonal and paracrine system. Its activation causes sodium and water retention and thereby contributes to edema formation. Blockade of the conversion of angiotensin I to AII and blockade of the AII receptors enhance sodium and water excretion and reduce many forms of edema. AII that enters the systemic circulation stimulates the production of aldosterone by the zona glomerulosa of the adrenal cortex. Aldosterone in turn enhances sodium reabsorption (and potassium excretion) by the collecting tubule, further favoring edema formation. Blockade of the action of aldosterone by spironolactone or eplerenone (aldosterone antagonists) or by amiloride (a blocker of epithelial sodium channels) often induces a moderate diuresis in edematous states.

ARGININE VASOPRESSIN

(See also Chap. 381) The secretion of arginine vasopressin (AVP) by the posterior pituitary gland occurs in response to increased intracellular osmolar concentration; by stimulating V₂ receptors, AVP increases the reabsorption of free water in the distal tubules and collecting ducts of the kidneys, thereby increasing total-body water. Circulating AVP is elevated in many patients with heart failure secondary to a nonosmotic stimulus associated with decreased effective arterial volume and reduced compliance of the left atrium. Such patients fail to show the normal reduction of AVP with a reduction of osmolality, contributing to edema formation and hyponatremia.

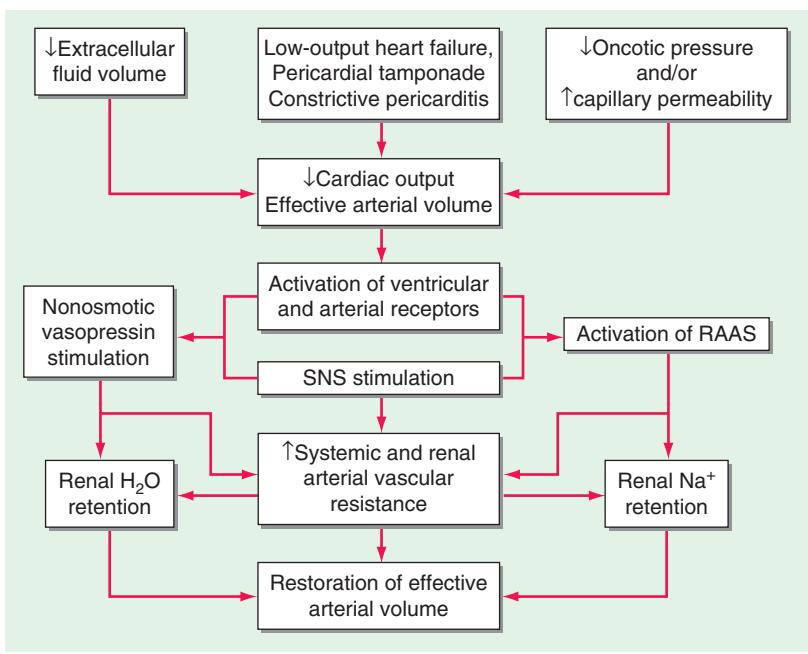
41

Edema

Joseph Loscalzo

PLASMA AND INTERSTITIAL FLUID EXCHANGE

Approximately two-thirds of total body water is intracellular and one-third is extracellular. One-fourth of the latter is in the plasma, and the remainder comprises the interstitial fluid. Edema represents an excess of interstitial fluid that has become evident clinically.



A

closely related natriuretic peptide (pre-pro-hormone brain natriuretic peptide [BNP]) is stored primarily in ventricular myocytes and is released when ventricular diastolic pressure rises. Released ANP and BNP (which is derived from its precursor) bind to the natriuretic receptor-A, which causes (1) excretion of sodium and water by augmenting glomerular filtration rate, inhibiting sodium reabsorption in the proximal tubule, and inhibiting release of renin and aldosterone; and (2) dilation of arterioles and venules by antagonizing the vasoconstrictor actions of AII, AVP, and sympathetic stimulation. Thus, elevated levels of natriuretic peptides have the capacity to oppose sodium retention in hypervolemic and edematous states.

Although circulating levels of ANP and BNP are elevated in heart failure and in cirrhosis with ascites, these natriuretic peptides are not sufficiently potent to prevent edema formation. Indeed, in edematous states, resistance to the actions of natriuretic peptides may be increased, further reducing their effectiveness.

Further discussion of the control of sodium and water balance is found in [Chap. S1](#).

CLINICAL CAUSES OF EDEMA

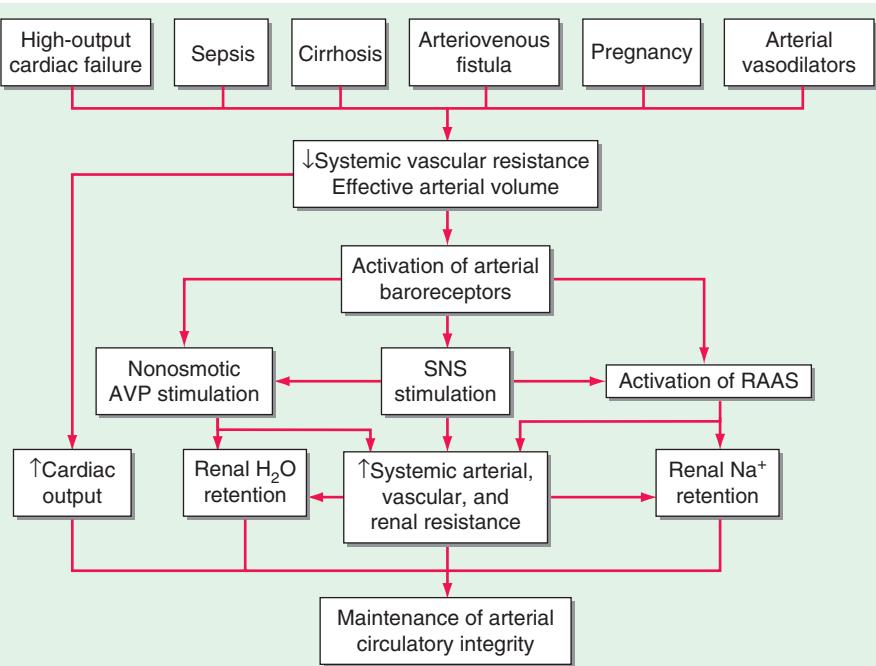
A weight gain of several kilograms usually precedes overt manifestations of generalized edema. *Anasarca* refers to gross, generalized edema. *Ascites* ([Chap. 50](#)) and *hydrothorax* refer to accumulation of excess fluid in the peritoneal and pleural cavities, respectively, and are considered special forms of edema.

Edema is recognized by the persistence of an indentation of the skin after pressure known as “pitting” edema. In its more subtle form, edema may be detected by noting that after the stethoscope is removed from the chest wall, the rim of the bell leaves an indentation on the skin of the chest for a few minutes. Edema may be present when the ring on a finger fits more snugly than in the past or when a patient complains of difficulty putting on shoes, particularly in the evening. Edema may also be recognized by puffiness of the face, which is most readily apparent in the periorbital areas owing to relative tissue laxity.

GENERALIZED EDEMA

The differences among the major causes of generalized edema are shown in [Table 41-1](#). Cardiac, renal, hepatic, or nutritional disorders are responsible for a large majority of patients with generalized edema. Consequently, the differential diagnosis of generalized edema should be directed toward identifying or excluding these several conditions.

Heart Failure ([See also Chap. 257](#)) In heart failure, the impaired systolic emptying of the ventricle(s) and/or the impairment of ventricular relaxation promotes an accumulation of blood in the venous circulation at the expense of the effective arterial volume. In addition, the activation of the sympathetic nervous system and the RAAS (see above) acts in concert to cause renal vasoconstriction and reduction of glomerular filtration and salt and water retention. Sodium and water retention continue, and the increment in blood volume accumulates in



B

FIGURE 41-1 Clinical conditions in which a decrease in cardiac output (A) and systemic vascular resistance (B) cause arterial underfilling with resulting neurohumoral activation and renal sodium and water retention. In addition to activating the neurohumoral axis, adrenergic stimulation causes renal vasoconstriction and enhances sodium and fluid transport by the proximal tubule epithelium. AVP, arginine vasopressin; RAAS, renin-angiotensin aldosterone system; SNS, sympathetic nervous system. (From Annals of Internal Medicine, RW Schrier: Body fluid volume regulation in health and disease: A unifying hypothesis. 113(2):155-159, 1990. Copyright © 1990, American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.)

ENDOTHELIN-1

This potent peptide vasoconstrictor is released by endothelial cells. Its concentration in the plasma is elevated in patients with severe heart failure and contributes to renal vasoconstriction, sodium retention, and edema.

NATRIURETIC PEPTIDES

Atrial distention causes release into the circulation of atrial natriuretic peptide (ANP), a polypeptide. A high-molecular-weight precursor of ANP is stored in secretory granules within atrial myocytes. A

TABLE 41-1 Principal Causes of Generalized Edema: History, Physical Examination, and Laboratory Findings

ORGAN SYSTEM	HISTORY	PHYSICAL EXAMINATION	LABORATORY FINDINGS
Cardiac	Dyspnea with exertion prominent—often associated with orthopnea—or paroxysmal nocturnal dyspnea	Elevated jugular venous pressure, ventricular (S_3) gallop; occasionally with displaced or dyskinetic apical pulse; peripheral cyanosis, cool extremities, small pulse pressure when severe	Elevated urea nitrogen-to-creatinine ratio common; serum sodium often diminished; elevated natriuretic peptides
Hepatic	Dyspnea uncommon, except if associated with significant degree of ascites; most often a history of ethanol abuse	Frequently associated with ascites; jugular venous pressure normal or low; blood pressure lower than in renal or cardiac disease; one or more additional signs of chronic liver disease (jaundice, palmar erythema, Dupuytren's contracture, spider angioma, male gynecomastia; asterixis and other signs of encephalopathy) may be present	If severe, reductions in serum albumin, cholesterol, other hepatic proteins (transferrin, fibrinogen); liver enzymes elevated, depending on the cause and acuity of liver injury; tendency toward hypokalemia, respiratory alkalosis; macrocytosis from folate deficiency
Renal (CRF)	Usually chronic: may be associated with uremic signs and symptoms, including decreased appetite, altered (metallic or fishy) taste, altered sleep pattern, difficulty concentrating, restless legs, or myoclonus; dyspnea can be present, but generally less prominent than in heart failure	Elevated blood pressure; hypertensive retinopathy; nitrogenous fetor; pericardial friction rub in advanced cases with uremia	Elevation of serum creatinine and cystatin C; albuminuria; hyperkalemia, metabolic acidosis, hyperphosphatemia, hypocalcemia, anemia (usually normocytic)
Renal (NS)	Childhood diabetes mellitus; plasma cell dyscrasias	Periorbital edema; hypertension	Proteinuria (≥ 3.5 g/d); hypoalbuminemia; hypercholesterolemia; microscopic hematuria

Abbreviations: CRF, chronic renal failure; NS, nephrotic syndrome.

Source: Reproduced with permission from GM Chertow, in E Braunwald, L Goldman (eds): Approach to the patient with edema, in Primary Cardiology, 2nd ed. Philadelphia, Saunders, 2003.

the venous circulation, raising venous and intracapillary pressure and resulting in edema (Fig. 41-1).

The presence of overt cardiac disease, as manifested by cardiac enlargement and/or ventricular hypertrophy, together with clinical evidence of cardiac failure, such as dyspnea, basilar rales, venous distension, and hepatomegaly, usually indicates that edema results from heart failure. Noninvasive tests such as electrocardiography, echocardiography, and measurements of BNP (or N-terminal proBNP [NT-proBNP]) are helpful in establishing the diagnosis of heart disease. The edema of heart failure typically occurs in the dependent portions of the body.

Edema of Renal Disease (See also Chap. 314) The edema that occurs during the acute phase of glomerulonephritis is characteristically associated with hematuria, proteinuria, and hypertension. In most instances, the edema results from primary retention of sodium and water by the kidneys owing to renal dysfunction. This state differs from most forms of heart failure in that it is characterized by a normal (or sometimes even increased) cardiac output. Patients with *chronic* renal failure may also develop edema due to primary renal retention of sodium and water.

Nephrotic Syndrome and Other Hypoalbuminemic States The primary alteration in the nephrotic syndrome is a diminished colloid oncotic pressure due to losses of large quantities (≥ 3.5 g/d) of protein into the urine and hypoalbuminemia (< 3.0 g/dL). As a result of the reduced colloid osmotic pressure, the sodium and water that are retained cannot be confined within the vascular compartment, and total and effective arterial blood volumes decline. This process initiates the edema-forming sequence of events described above, including activation of the RAAS. The nephrotic syndrome may occur during the course of a variety of kidney diseases, including glomerulonephritis, diabetic glomerulosclerosis, and hypersensitivity reactions. The edema is diffuse, symmetric, and most prominent in the dependent areas; periorbital edema is most prominent in the morning.

Hepatic Cirrhosis (See also Chap. 344) This condition is characterized, in part, by hepatic venous outflow obstruction, which in turn expands the splanchnic blood volume, and hepatic lymph formation. Intrahepatic hypertension acts as a stimulus for renal sodium retention and causes a reduction of effective arterial blood volume. These alterations are frequently complicated by hypoalbuminemia secondary to reduced hepatic synthesis, as well as peripheral arterial vasodilation. These effects reduce the effective arterial blood volume, leading to activation

of the sodium- and water-retaining mechanisms described above (Fig. 41-1B). The concentration of circulating aldosterone often is elevated by the failure of the liver to metabolize this hormone. Initially, the excess interstitial fluid is localized preferentially proximal (upstream) to the congested portal venous system, causing ascites (Chap. 50). In later stages, particularly when there is severe hypoalbuminemia, peripheral edema may develop. A sizable accumulation of ascitic fluid may increase intraabdominal pressure and impede venous return from the lower extremities and contribute to the accumulation of the edema.

Drug-Induced Edema A large number of widely used drugs can cause edema (Table 41-2). Mechanisms include renal vasoconstriction

TABLE 41-2 Drugs Associated with Edema Formation

Nonsteroidal anti-inflammatory drugs
Antihypertensive agents
Direct arterial/arteriolar vasodilators
Hydralazine
Clonidine
Methyldopa
Guanethidine
Minoxidil
Calcium channel antagonists
α -Adrenergic antagonists
Thiazolidinediones
Steroid hormones
Glucocorticoids
Anabolic steroids
Estrogens
Progesterins
Cyclosporine
Growth hormone
Immunotherapies
Interleukin 2
OKT3 monoclonal antibody

Source: Reproduced with permission from GM Chertow, in E Braunwald, L Goldman (eds): Approach to the patient with edema, in Primary Cardiology, 2nd ed. Philadelphia, Saunders, 2003.

(nonsteroidal anti-inflammatory drugs and cyclosporine), arteriolar dilation (vasodilators), augmented renal sodium reabsorption (steroid hormones), and capillary damage.

Edema of Nutritional Origin A diet grossly deficient in calories and particularly in protein over a prolonged period may produce hypoproteinemia and edema. The latter may be intensified by the development of beriberi heart disease, which also is of nutritional origin, in which multiple peripheral arteriovenous fistulae result in reduced effective systemic perfusion and effective arterial blood volume, thereby enhancing edema formation (Chap. 333) (Fig. 41-1B). Edema develops or becomes intensified when famished subjects are first provided with an adequate diet. The ingestion of more food may increase the quantity of sodium ingested, which is then retained along with water. So-called refeeding edema also may be linked to increased release of insulin, which directly increases tubular sodium reabsorption. In addition to hypoalbuminemia, hypokalemia and caloric deficits may be involved in the edema of starvation.

■ LOCALIZED EDEMA

In thrombophlebitis, varicose veins, and primary venous valve failure, the hydrostatic pressure in the capillary bed upstream (proximal) of the obstruction increases so that an abnormal quantity of fluid is transferred from the vascular to the interstitial space, which may give rise to localized edema. The latter may also occur in lymphatic obstruction caused by chronic lymphangitis, resection of regional lymph nodes, filariasis, and genetic (frequently called primary) lymphedema. The latter is particularly intractable because restriction of lymphatic flow results in both an increase in intracapillary pressure and increased protein concentration in the interstitial fluid, which act in concert to aggravate fluid retention.

Other Causes of Edema These causes include hypothyroidism (myxedema) due to deposition of hyaluronic acid; hyperthyroidism (pretibial myxedema secondary to Graves' disease), in which edema is typically nonpitting and, in Graves' disease, exogenous hypercortisolism; pregnancy; and administration of estrogens and vasodilators, particularly dihydropyridines such as nifedipine.

■ DISTRIBUTION OF EDEMA

The distribution of edema is an important guide to its cause. Edema associated with heart failure tends to be more extensive in the legs and to be accentuated in the evening, a feature also determined largely by posture. When patients with heart failure are confined to bed, edema may be most prominent in the presacral region.

Edema resulting from hypoproteinemia, as occurs in the nephrotic syndrome, characteristically is generalized, but it is especially evident in the very soft tissues of the eyelids and face and tends to be most pronounced in the morning owing to the recumbent posture assumed during the night. Less common causes of facial edema include trichinosis, allergic reactions, and myxedema. Edema limited to one leg or to one or both arms is usually the result of venous and/or lymphatic obstruction. Unilateral paralysis reduces lymphatic and venous drainage on the affected side and may also be responsible for unilateral edema. In patients with obstruction of the superior vena cava, edema is confined to the face, neck, and upper extremities in which the venous pressure is elevated compared with that in the lower extremities.

APPROACH TO THE PATIENT

Edema

An important first question is whether the edema is localized or generalized. If it is localized, the local phenomena that may be responsible should be identified. If the edema is generalized, one should determine if there is serious hypoalbuminemia, e.g., serum albumin <3.0 g/dL. If so, the history, physical examination, urinalysis, and other laboratory data will help evaluate the question of cirrhosis, severe malnutrition, or the nephrotic syndrome as the

underlying disorder. If hypoalbuminemia is not present, it should be determined if there is evidence of heart failure severe enough to promote generalized edema. Finally, it should be ascertained as to whether or not the patient has an adequate urine output or if there is significant oliguria or anuria. **These abnormalities are discussed in Chaps. 52, 310, and 311.**

■ FURTHER READING

- CLARK AL, CLELAND JG: Causes and treatment of oedema in patients with heart failure. *Nature Rev Cardiol* 10:156, 2013.
- DAMMAN K et al: Congestion in chronic systolic heart failure is related to renal dysfunction and increased mortality. *Eur J Heart Fail* 12:974, 2010.
- FERRELL RE et al: *GJC2* missense mutations cause human lymphedema. *Am J Hum Genet* 86:943, 2010.
- FRISON S et al: Omitting edema measurement: How much acute malnutrition are we missing? *Am J Clin Nutr* 102:1176, 2015.
- LEVICK JR, MICHEL CC: Microvascular fluid exchange and the revised Starling principle. *Cardiovascular Res* 87:198, 2010.
- TELINIUS N, HJORTDAL VE: Role of the lymphatic vasculature in cardiovascular medicine. *Heart* 105:1777, 2019.

42

Approach to the Patient with a Heart Murmur

Patrick T. O'Gara, Joseph Loscalzo



The differential diagnosis of a heart murmur begins with a careful assessment of its major attributes and response to bedside maneuvers. The history, clinical context, and associated physical examination findings provide additional clues to help establish the significance of a heart murmur. Accurate bedside identification of a heart murmur can inform decisions regarding the indications for noninvasive testing and the need for referral to a cardiovascular specialist. Preliminary discussions can be held with the patient regarding antibiotic or rheumatic fever prophylaxis, the need to restrict various forms of physical activity, and the potential role for family screening.

Heart murmurs are caused by audible vibrations that are due to increased turbulence from accelerated blood flow through normal or abnormal orifices; flow through a narrowed or irregular orifice into a dilated vessel or chamber; or backward flow through an incompetent valve, ventricular septal defect, or patent ductus arteriosus. They traditionally are defined by their timing within the cardiac cycle (Fig. 42-1). *Systolic murmurs* begin with or after the first heart sound (S_1) and terminate at or before the component (A_2 or P_2) of the second heart sound (S_2) that corresponds to their site of origin (left or right, respectively). *Diastolic murmurs* begin with or after the associated component of S_2 and end at or before the subsequent S_1 . *Continuous murmurs* are not confined to either phase of the cardiac cycle but instead begin in early systole and proceed through S_2 into all or part of diastole. The accurate timing of heart murmurs is the first step in their identification. The distinction between S_1 and S_2 , and therefore systole and diastole, is usually a straightforward process but can be difficult in the setting of a tachyarrhythmia, in which case the heart sounds can be distinguished by simultaneous palpation of the carotid upstroke, which should closely follow S_1 .

Duration and Character The duration of a heart murmur depends on the length of time over which a pressure difference exists between two cardiac chambers, the left ventricle and the aorta, the right ventricle and the pulmonary artery, or the great vessels. The magnitude and

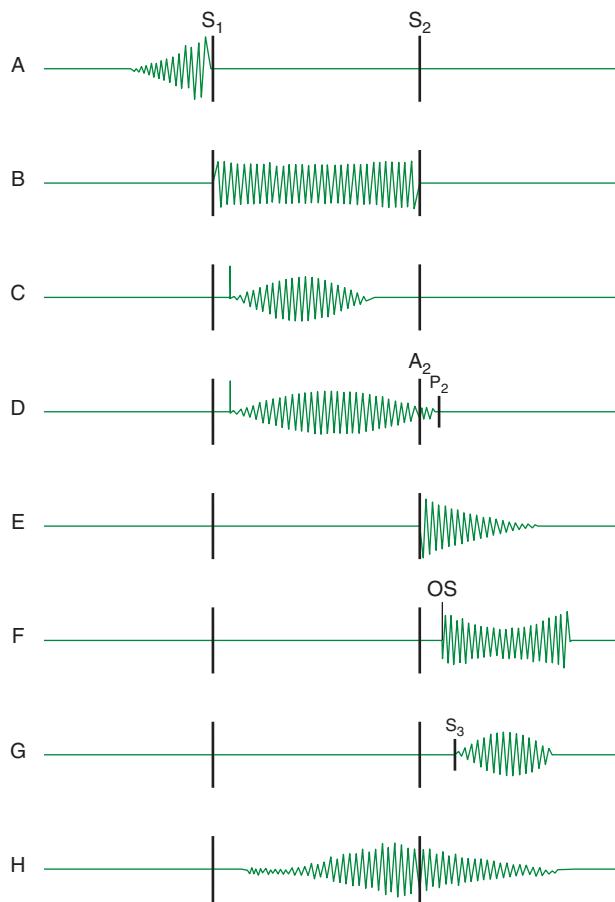


FIGURE 42-1 Diagram depicting principal heart murmurs. **A.** Presystolic murmur of mitral or tricuspid stenosis. **B.** Holosystolic (pansystolic) murmur of mitral or tricuspid regurgitation or of ventricular septal defect. **C.** Aortic ejection murmur beginning with an ejection click and fading before the second heart sound. **D.** Systolic murmur in pulmonic stenosis spilling through the aortic second sound, pulmonic valve closure being delayed. **E.** Aortic or pulmonary diastolic murmur. **F.** Long diastolic murmur of mitral stenosis after the opening snap (OS). **G.** Short mid-diastolic inflow murmur after a third heart sound. **H.** Continuous murmur of patent ductus arteriosus. (Courtesy of Antony and Julie Wood.)

variability of this pressure difference, coupled with the geometry and compliance of the involved chambers or vessels, dictate the velocity of flow; the degree of turbulence; and the resulting frequency, configuration, and intensity of the murmur. The diastolic murmur of chronic aortic regurgitation (AR) is a blowing, high-frequency event, whereas the murmur of mitral stenosis (MS), indicative of the left atrial-left ventricular diastolic pressure gradient, is a low-frequency event, heard as a rumbling sound with the bell of the stethoscope. The frequency components of a heart murmur may vary at different sites of auscultation. The coarse systolic murmur of aortic stenosis (AS) may sound higher pitched and more acoustically pure at the apex, a phenomenon eponymously referred to as the *Gallavardin effect*. Some murmurs may have a distinct or unusual quality, such as the “honking” sound appreciated in some patients with mitral regurgitation (MR) due to mitral valve prolapse (MVP).

The configuration of a heart murmur may be described as crescendo, decrescendo, crescendo-decrescendo, or plateau. The decrescendo configuration of the murmur of chronic AR (Fig. 42-1E) can be understood in terms of the progressive decline in the diastolic pressure gradient between the aorta and the left ventricle. The crescendo-decrescendo configuration of the murmur of AS reflects the changes in the systolic pressure gradient between the left ventricle and the aorta as ejection occurs, whereas the plateau configuration of the murmur of chronic MR (Fig. 42-1B) is consistent with the large and nearly constant pressure difference between the left ventricle and the left atrium.

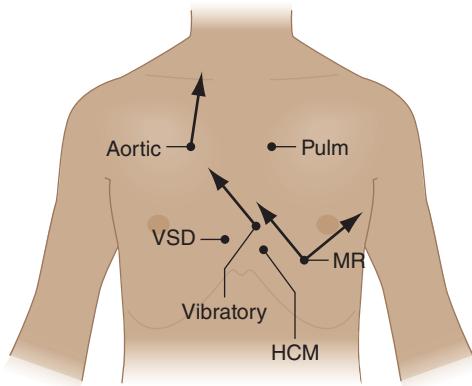


FIGURE 42-2 Maximal intensity and radiation of six isolated systolic murmurs. Aortic, aortic stenosis; HCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; Pulm, pulmonary stenosis; VSD, ventricular septal defect. (From JB Barlow: Perspectives on the Mitral Valve. Philadelphia, FA Davis, 1987, p 140.)

Intensity The intensity of a heart murmur is graded on a scale of 1–6 (or I–VI). A grade 1 murmur is very soft and is heard only with great effort. A grade 2 murmur is easily heard but not particularly loud. A grade 3 murmur is loud but is not accompanied by a palpable thrill over the site of maximal intensity. A grade 4 murmur is very loud and accompanied by a thrill. A grade 5 murmur is loud enough to be heard with only the edge of the stethoscope touching the chest, whereas a grade 6 murmur is loud enough to be heard with the stethoscope slightly off the chest. Murmurs of grade 3 or greater intensity usually signify important structural heart disease and indicate high blood flow velocity at the site of murmur production. Small, restrictive ventricular septal defects (VSDs), for example, are accompanied by loud, usually grade 4 or greater, systolic murmurs as blood is ejected at high velocity from the left ventricle to the right ventricle. Low-velocity events, such as left-to-right shunting across an atrial septal defect (ASD), are usually silent. The intensity of a heart murmur may be diminished by any process that increases the distance between the intracardiac source and the stethoscope on the chest wall, such as obesity, obstructive lung disease, or a large pericardial effusion. The intensity of a murmur also may be misleadingly soft when cardiac output is reduced significantly or when the pressure gradient between the involved cardiac structures is low.

Location and Radiation Recognition of the location and radiation of the murmur helps facilitate its accurate identification (Fig. 42-2). Adventitious sounds, such as a systolic click or diastolic snap, or abnormalities of S_1 or S_2 , may provide additional clues. Careful attention to the characteristics of the murmur and other heart sounds during the respiratory cycle and the performance of simple bedside maneuvers complete the auscultatory examination. These features, along with recommendations for further testing, are discussed below in the context of specific systolic, diastolic, and continuous heart murmurs (Table 42-1).

■ SYSTOLIC HEART MURMURS

Early Systolic Murmurs Early systolic murmurs begin with S_1 and extend for a variable period, ending well before S_2 . Their causes are relatively few. Acute, severe MR into a normal-sized, relatively non-compliant left atrium results in an early, decrescendo systolic murmur best heard at or just medial to the apical impulse. These characteristics reflect the progressive attenuation of the pressure gradient between the left ventricle and the left atrium during systole owing to the rapid rise in left atrial pressure caused by the sudden volume load into an unprepared, noncompliant chamber, and contrast sharply with the auscultatory features of chronic MR. Clinical settings in which acute, severe MR occur include (1) papillary muscle rupture complicating acute myocardial infarction (MI) (Chap. 275), (2) rupture of chordae tendineae in the setting of myxomatous mitral valve disease (MVP, Chap. 265), (3) infective endocarditis (Chap. 128), and (4) blunt chest wall trauma.

TABLE 42-1 Principal Causes of Heart Murmurs**Systolic Murmurs**

Early systolic	
Mitral	
Acute MR	
VSD	
Muscular	
Nonrestrictive with pulmonary hypertension	
Tricuspid	
TR with normal pulmonary artery pressure	
Midsystolic	
Aortic	
Obstructive	
Supravalvular-supravalvular AS, coarctation of the aorta	
Valvular-AS and aortic sclerosis	
Subvalvular-discrete, tunnel or HOCM	
Increased flow, hyperkinetic states, AR, complete heart block	
Dilation of ascending aorta, atheroma, aortitis	
Pulmonary	
Obstructive	
Supravalvular-pulmonary artery stenosis	
Valvular-pulmonic valve stenosis	
Subvalvular-infundibular stenosis (dynamic)	
Increased flow, hyperkinetic states, left-to-right shunt (e.g., ASD)	
Dilation of pulmonary artery	
Late systolic	
Mitral	
MVP, acute myocardial ischemia	
Tricuspid	
TVP	
Holosystolic	
Atrioventricular valve regurgitation (MR, TR)	
Left-to-right shunt at ventricular level (VSD)	

Early Diastolic Murmurs

AR	
Valvular: congenital (bicuspid valve), rheumatic deformity, endocarditis, prolapse, trauma, post-valvotomy	
Dilation of valve ring: aorta dissection, annuloaortic ectasia, medial degeneration, hypertension, ankylosing spondylitis	
Widening of commissures: syphilis	
Pulmonic regurgitation	
Valvular: post-valvotomy, endocarditis, rheumatic fever, carcinoid	
Dilation of valve ring: pulmonary hypertension; Marfan syndrome	
Congenital: isolated or associated with tetralogy of Fallot, VSD, pulmonic stenosis	

Mid-Diastolic Murmurs

Mitral	
MS	
Carey-Coombs murmur (mid-diastolic apical murmur in acute rheumatic fever)	
Increased flow across nonstenotic mitral valve (e.g., MR, VSD, PDA, high-output states, and complete heart block)	
Tricuspid	
Tricuspid stenosis	
Increased flow across nonstenotic tricuspid valve (e.g., TR, ASD, and anomalous pulmonary venous return)	
Left and right atrial tumors (myxoma)	
Severe AR (Austin Flint murmur)	

Continuous Murmurs

Patent ductus arteriosus	Proximal coronary artery stenosis
Coronary AV fistula	Mammary souffle of pregnancy
Ruptured sinus of Valsalva aneurysm	Pulmonary artery branch stenosis
Aortic septal defect	Bronchial collateral circulation
Cervical venous hum	Small (restrictive) ASD with MS
Anomalous left coronary artery	Intercostal AV fistula

Abbreviations: AR, aortic regurgitation; AS, aortic stenosis; ASD, atrial septal defect; AV, arteriovenous; HOCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; MS, mitral stenosis; MVP, mitral valve prolapse; PDA, patent ductus arteriosus; TR, tricuspid regurgitation; TVP, tricuspid valve prolapse; VSD, ventricular septal defect.

Source: E Braunwald, JK Perloff, in D Zipes et al (eds): *Braunwald's Heart Disease*, 7th ed. Philadelphia, Elsevier, 2005; PJ Norton, RA O'Rourke, in E Braunwald, L Goldman (eds): *Primary Cardiology*, 2nd ed. Philadelphia, Elsevier, 2003.

Acute, severe MR from papillary muscle rupture usually accompanies an inferior, posterior, or lateral MI and occurs 2–7 days after presentation. It often is signaled by chest pain, hypotension, and pulmonary edema, but a murmur may be absent in up to 50% of cases. The posteromedial papillary muscle is involved 6–10 times more frequently than the anterolateral papillary muscle. The murmur is to be distinguished from that associated with post-MI ventricular septal rupture, which is accompanied by a systolic thrill at the left sternal border in nearly all patients and is holosystolic in duration. A new heart murmur after an MI is an indication for transthoracic echocardiography (TTE) (Chap. 241), which allows bedside delineation of its etiology and pathophysiologic significance. The distinction between acute MR and ventricular septal rupture also can be achieved with right-sided heart catheterization, sequential determination of oxygen saturations, and analysis of the pressure waveforms (tall *v* wave in the pulmonary artery wedge pressure in MR). Post-MI mechanical complications of this nature mandate aggressive medical stabilization and prompt referral for surgical repair.

Spontaneous chordal rupture can complicate the course of myxomatous mitral valve disease (MVP) and result in new-onset or “acute on chronic” severe MR. MVP may occur as an isolated phenomenon, or the lesion may be part of a more generalized connective tissue disorder as seen, for example, in patients with Marfan syndrome. Acute, severe MR as a consequence of infective endocarditis results from destruction of leaflet tissue, chordal rupture, or both. Blunt chest wall trauma is usually self-evident but may be disarmingly trivial; it can result in papillary muscle contusion and rupture, chordal detachment, or leaflet avulsion. TTE is indicated in all cases of suspected acute, severe MR to define its mechanism and severity, delineate left ventricular size and systolic function, and provide an assessment of suitability for primary valve repair.

A congenital, small muscular VSD (Chap. 269) may be associated with an early systolic murmur. The defect closes progressively during septal contraction, and thus the murmur is confined to early systole. It is localized to the left sternal border (Fig. 42-2) and is usually of grade 4 or 5 intensity. Signs of pulmonary hypertension or left ventricular volume overload are absent. Anatomically large and uncorrected VSDs, which usually involve the membranous portion of the septum, may lead to pulmonary hypertension. The murmur associated with the left-to-right shunt, which earlier may have been holosystolic, becomes limited to the first portion of systole as the elevated pulmonary vascular resistance leads to an abrupt rise in right ventricular pressure and an attenuation of the interventricular pressure gradient during the remainder of the cardiac cycle. In such instances, signs of pulmonary hypertension (right ventricular lift, loud and single or closely split S₂) may predominate. The murmur is best heard along the left sternal border but is softer. Suspicion of a VSD is an indication for TTE.

Tricuspid regurgitation (TR) with normal pulmonary artery pressures, as may occur with infective endocarditis, may produce an early systolic murmur. The murmur is soft (grade 1 or 2), is best heard at the lower left sternal border, and may increase in intensity with inspiration (Carvallo's sign). Regurgitant *c-v* waves may be visible in the jugular venous pulse. TR in this setting is not associated with signs of right heart failure, such as ascites or lower extremity edema.

Midsystolic Murmurs Midsystolic murmurs begin at a short interval after S₁, end before S₂ (Fig. 42-1C) and are usually crescendo-decrescendo in configuration. AS is the most common cause of a midsystolic murmur in an adult. The murmur of AS is usually loudest to the right of the sternum in the second intercostal space (aortic area, Fig. 42-2) and radiates into the carotids. Transmission of the midsystolic murmur to the apex, where it becomes higher-pitched, is common (Gallavardin effect; see above).

Differentiation of this apical systolic murmur from MR can be difficult. The murmur of AS will increase in intensity or become louder, in the beat after a premature beat, whereas the murmur of MR will have constant intensity from beat to beat. The intensity of the AS murmur also varies directly with the cardiac output. With a normal cardiac output, a systolic thrill at the second right intercostal space and a

grade 4 or higher murmur suggest severe AS. The murmur is softer in the setting of heart failure and low cardiac output. Other auscultatory findings of severe AS include a soft or absent A_2 , paradoxical splitting of S_2 , an apical S_4 , and a late-peaking systolic murmur. In children, adolescents, and young adults with congenital valvular AS, an early ejection sound (click) is usually audible, more often along the left sternal border than at the base. Its presence signifies a flexible, noncalcified bicuspid valve (or one of its variants) and localizes the left ventricular outflow obstruction to the valvular (rather than sub- or supravalvular) level.

Assessment of the volume and rate of rise of the carotid pulse can provide additional information. A small and delayed upstroke (*parvus et tardus*) is consistent with severe AS. The carotid pulse examination is less discriminatory, however, in older patients with stiffened arteries. The electrocardiogram (ECG) shows signs of left ventricular hypertrophy (LVH) as the severity of the stenosis increases. TTE is indicated to assess the anatomic features of the aortic valve, the severity of the stenosis, left ventricular size, wall thickness and function, and the size and contour of the aortic root and proximal ascending aorta.

The obstructive form of hypertrophic cardiomyopathy (HOCM) is associated with a midsystolic murmur that is usually loudest along the left sternal border or between the left lower sternal border and the apex (Chap. 259, Fig. 42-2). The murmur is produced by both dynamic left ventricular outflow tract obstruction and MR, and thus, its configuration is a hybrid between ejection and regurgitant phenomena. The intensity of the murmur may vary from beat to beat and after provocative maneuvers but usually does not exceed grade 3. The murmur classically will increase in intensity with maneuvers that result in increasing degrees of outflow tract obstruction, such as a reduction in preload or afterload (Valsalva, standing, vasodilators), or with an augmentation of contractility (inotropic stimulation). Maneuvers or medications that increase preload (squatting, passive leg raising, volume administration) or afterload (squatting, vasopressors) or that reduce contractility (β -adrenoreceptor blockers) decrease the intensity of the murmur. In rare patients, there may be reversed splitting of S_2 . A sustained left ventricular apical impulse and an S_4 may be appreciated. In contrast to AS, the carotid upstroke is rapid and of normal volume. Rarely, it is bisferiens or bifid in contour (see Fig. 239-2D) due to midsystolic closure of the aortic valve. LVH is present on the ECG, and the diagnosis is confirmed by TTE. Although the systolic murmur associated with MVP behaves similarly to that due to HOCM in response to the Valsalva maneuver and to standing/squatting (Fig. 42-3), these two lesions can be distinguished on the basis of their associated findings, such as the presence of LVH in HOCM or a nonejection click in MVP.

The midsystolic, crescendo-decrescendo murmur of congenital pulmonic stenosis (PS; Chap. 269) is best appreciated in the second and third left intercostal spaces (pulmonic area) (Figs. 42-2 and 42-4). The duration of the murmur lengthens and the intensity of P_2 diminishes with increasing degrees of valvular stenosis (Fig. 42-1D). An early ejection sound, the intensity of which *decreases* with inspiration, is heard in younger patients. A parasternal lift and ECG evidence of right ventricular hypertrophy indicate severe pressure overload. If obtained, the chest x-ray may show poststenotic dilation of the main pulmonary artery. TTE is recommended for complete characterization.

Significant left-to-right intracardiac shunting due to an ASD (Chap. 269) leads to an increase in pulmonary blood flow and a grade 2–3 midsystolic murmur at the middle to upper left sternal border attributed to increased flow rates across the pulmonic valve with fixed splitting of S_2 . Ostium secundum ASDs are the most common cause of these shunts in adults. Features suggestive of a primum ASD include the coexistence of MR due to a cleft anterior mitral valve leaflet and left axis deviation of the QRS complex on the ECG. With sinus venosus ASDs, the left-to-right shunt is usually not large enough to result in a systolic murmur, although the ECG may show abnormalities of sinus node function. A grade 2 or 3 midsystolic murmur may also be heard best at the upper left sternal border in patients with idiopathic dilation of the pulmonary artery; a pulmonary ejection sound is also present in these patients. TTE is indicated to evaluate a grade 2 or 3 midsystolic murmur when there are other signs of cardiac disease.

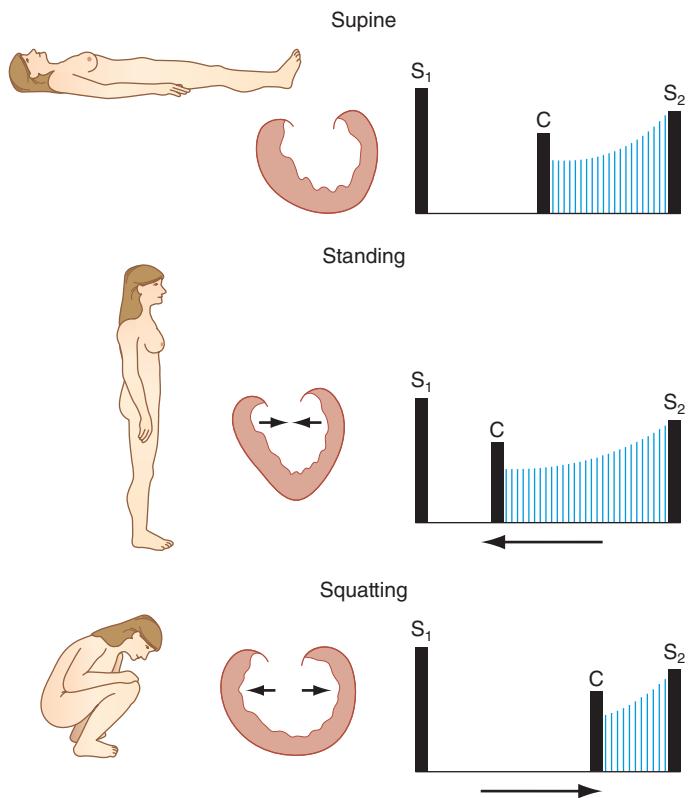
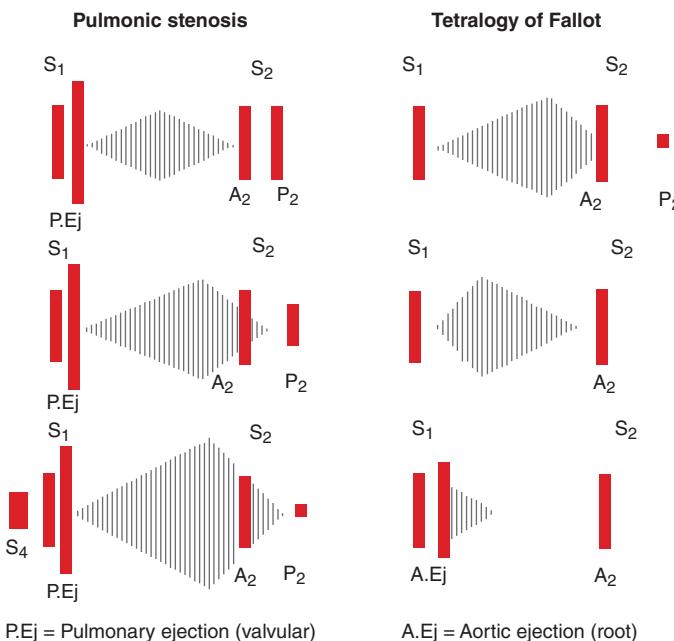


FIGURE 42-3 A midsystolic nonejection sound (C) occurs in mitral valve prolapse and is followed by a late systolic murmur that crescendos to the second heart sound (S_2). Standing decreases venous return; the heart becomes smaller; C moves closer to the first heart sound (S_1), and the mitral regurgitant murmur has an earlier onset. With prompt squatting, venous return and afterload increase; the heart becomes larger; C moves toward S_2 , and the duration of the murmur shortens. The systolic murmur of hypertrophic obstructive cardiomyopathy behaves similarly. (Reprinted with permission Examination of the Heart, Part IV: Auscultation of the Heart ©American Heart Association, Inc.)

An isolated grade 1 or 2 midsystolic murmur, heard in the absence of symptoms or signs of heart disease, is most often a benign finding for which no further evaluation, including TTE, is necessary. The most common example of a murmur of this type in an older adult patient is the crescendo-decrescendo murmur of aortic valve sclerosis, heard at the second right interspace (Fig. 42-2). Aortic sclerosis is defined as focal thickening and calcification of the aortic valve to a degree that does not interfere with leaflet opening. The carotid upstrokes are normal, and electrocardiographic LVH is not present. A grade 1 or 2 midsystolic murmur often can be heard at the left sternal border with pregnancy, hyperthyroidism, or anemia, physiologic states that are associated with accelerated blood flow. *Still's murmur* refers to a benign grade 2, vibratory or musical midsystolic murmur at the mid or lower left sternal border in normal children and adolescents, best heard in the supine position (Fig. 42-2).

Late Systolic Murmurs A late systolic murmur that is best heard at the left ventricular apex is usually due to MVP (Chap. 265). Often, this murmur is introduced by one or more nonejection clicks. The radiation of the murmur can help identify the specific mitral leaflet involved in the process of prolapse or flail. The term *flail* refers to the movement made by an unsupported portion of the leaflet (usually the tip) after loss of its chordal attachment(s). With posterior leaflet prolapse or flail, the resultant jet of MR is directed anteriorly and medially, as a result of which the murmur radiates to the base of the heart and masquerades as AS. Anterior leaflet prolapse or flail results in a posteriorly directed MR jet that radiates to the axilla or left infrascapular region. Leaflet flail is associated with a murmur of grade 3 or 4 intensity that can be heard throughout the precordium in thin-chested patients. The presence of an S_3 or a short, rumbling mid-diastolic murmur due to enhanced flow signifies severe MR.



P.Ej = Pulmonary ejection (valvular)

A.Ej = Aortic ejection (root)

FIGURE 42-4 Left. In valvular pulmonic stenosis with intact ventricular septum, right ventricular systolic ejection becomes progressively longer, with increasing obstruction to flow. As a result, the murmur becomes longer and louder, enveloping the aortic component of the second heart sound (A₂). The pulmonic component (P₂) occurs later, and splitting becomes wider but more difficult to hear because A₂ is lost in the murmur and P₂ becomes progressively fainter and lower pitched. As the pulmonic gradient increases, the isometric contraction phase shortens until the pulmonic valve ejection sound fuses with the first heart sound (S₁). In severe pulmonic stenosis with concentric hypertrophy and decreasing right ventricular compliance, a fourth heart sound appears. **Right.** In tetralogy of Fallot with increasing obstruction at the pulmonic infundibular area, an increasing amount of right ventricular blood is shunted across the silent ventricular septal defect and flow across the obstructed outflow tract decreases. Therefore, with increasing obstruction, the murmur becomes shorter, earlier, and fainter. P₂ is absent in severe tetralogy of Fallot. A large aortic root receives almost all cardiac output from both ventricular chambers, and the aorta dilates and is accompanied by a root ejection sound that does not vary with respiration. (Reprinted with permission Examination of the Heart, Part IV: Auscultation of the Heart ©American Heart Association, Inc.)

Bedside maneuvers that decrease left ventricular preload, such as standing, will cause the click and murmur of MVP to move closer to the first heart sound, as leaflet prolapse occurs earlier in systole. Standing also causes the murmur to become louder and longer. With squatting, left ventricular preload and afterload are increased abruptly, leading to an increase in left ventricular volume, and the click and murmur move away from the first heart sound as leaflet prolapse is delayed; the murmur becomes softer and shorter in duration (Fig. 42-3). As noted above, these responses to standing and squatting are directionally similar to those observed in patients with HOCM.

A late, apical systolic murmur indicative of MR may be heard transiently in the setting of acute myocardial ischemia; it is due to apical tethering and malcoaptation of the leaflets in response to structural and functional changes of the ventricle and mitral annulus. The intensity of the murmur varies as a function of left ventricular afterload and will increase in the setting of hypertension. TTE is recommended for assessment of late systolic murmurs.

Holosystolic Murmurs (Figs. 42-1B and 42-5) Holosystolic murmurs begin with S₁ and continue through systole to S₂. They are usually indicative of chronic mitral or tricuspid valve regurgitation or a VSD and warrant TTE for further characterization. The holosystolic murmur of chronic MR is best heard at the left ventricular apex and radiates to the axilla (Fig. 42-2); it is usually high-pitched and plateau in configuration because of the wide difference between left ventricular and left atrial pressure throughout systole. In contrast to acute MR, left atrial compliance is normal or even increased in chronic MR. As a result, there is only a small increase in left atrial pressure for any increase in regurgitant volume.

Several conditions are associated with chronic MR and an apical holosystolic murmur, including rheumatic scarring of the leaflets, mitral annular calcification, postinfarction left ventricular remodeling, and severe left ventricular chamber enlargement in the setting of a dilated cardiomyopathy (Chap. 259). The severity of the MR is worsened by any contribution from apical displacement of the papillary muscles and leaflet tethering (remodeling). Because the mitral annulus is contiguous with the left atrial endocardium, gradual enlargement of the left atrium from chronic MR will result in further stretching of the annulus and more MR; thus, “MR begets MR.” Chronic severe MR results in enlargement and leftward displacement of the left ventricular

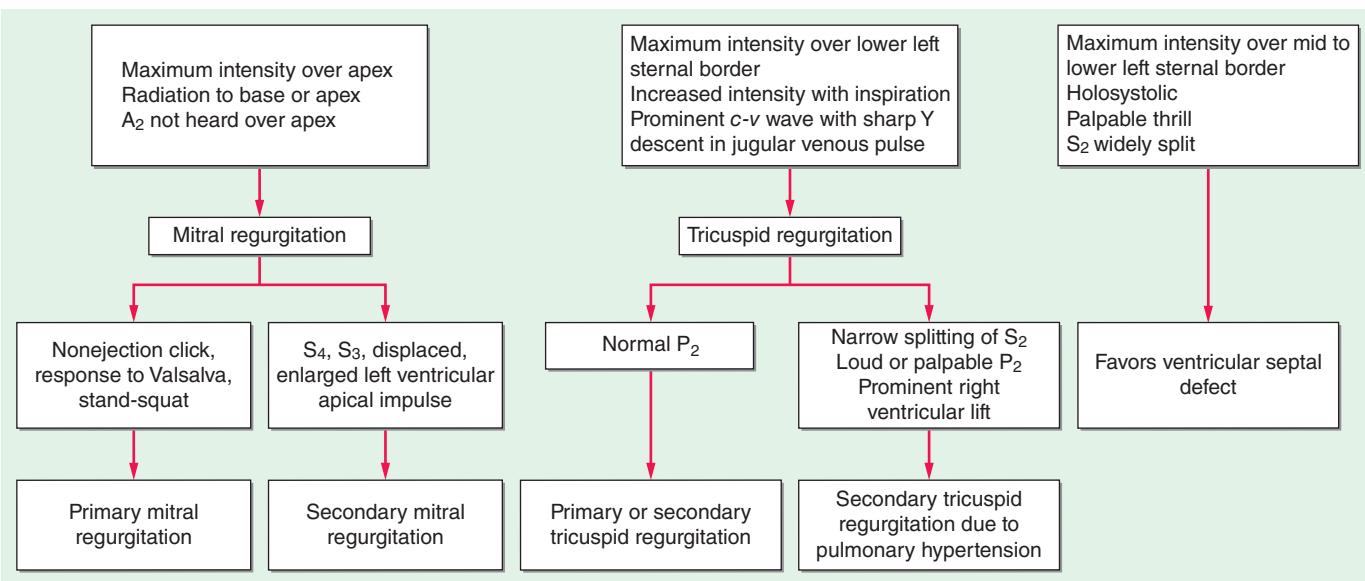


FIGURE 42-5 Differential diagnosis of a holosystolic murmur. The murmur of mitral regurgitation is best heard over the left ventricular apex. The radiation of the murmur depends on the direction in which the jet of mitral regurgitation enters into the left atrium. Differentiation of primary and secondary causes of mitral regurgitation is usually accomplished with transthoracic echocardiography, although the presence of a nonejection click and a mid-late apical systolic murmur, for example, can establish a bedside diagnosis of mitral valve prolapse (primary mitral regurgitation). Secondary mitral regurgitation can occur as a result of left ventricular remodeling. The murmur may be soft and difficult to hear. Other signs of left ventricular dysfunction may be present. Greater than 80% of the tricuspid regurgitation encountered clinically is due to a secondary cause. Severe pulmonary hypertension can be appreciated by a loud, single P₂. Primary tricuspid regurgitation may be present in the setting of pacemaker leads or in patients with carcinoid syndrome who usually have signs of liver involvement. A ventricular septal defect is usually manifested by a holosystolic murmur with a palpable thrill along the mid- to lower left sternal edge.

apex beat and, in some patients, a diastolic filling complex, as described previously (Fig. 42-1G).

The holosystolic murmur of chronic TR is generally softer than that of MR, is loudest at the left lower sternal border, and usually increases in intensity with inspiration (Carvallo's sign). Associated signs include *c-v* waves in the jugular venous pulse, an enlarged and pulsatile liver, ascites, and peripheral edema. The abnormal jugular venous waveforms are the predominant finding and seen very often in the absence of an audible murmur despite Doppler echocardiographic verification of TR. Causes of *primary* TR include myxomatous disease (prolapse), endocarditis, rheumatic disease, radiation, carcinoid, Ebstein's anomaly, leaflet trauma due to intracardiac device leads, or chordal detachment as a complication of right ventricular endomyocardial biopsy. TR is much more commonly a passive process that results secondarily from annular enlargement due to right ventricular dilation in the face of volume or pressure overload or adverse right ventricular remodeling.

The holosystolic murmur of a VSD is loudest at the mid- to lower-left sternal border (Fig. 42-2) and radiates widely. A thrill is present at the site of maximal intensity in the majority of patients. There is no change in the intensity of the murmur with inspiration. The intensity of the murmur varies as a function of the anatomic size of the defect. Small, restrictive VSDs, as exemplified by the *maladie de Roger*, create a very loud murmur due to the significant and sustained systolic pressure gradient between the left and right ventricles. With large defects, the ventricular pressures tend to equalize, shunt flow is balanced, and a murmur is not appreciated. The distinction between post-MI ventricular septal rupture and MR has been reviewed previously.

■ DIASTOLIC HEART MURMURS

Early Diastolic Murmurs (Fig. 42-1E) Chronic AR results in a high-pitched, blowing, decrescendo, early- to mid-diastolic murmur that begins after the aortic component of S₂ (A₂) and is best heard at the second right interspace and along the left sternal border. The murmur may be soft and difficult to hear unless auscultation is performed with the patient leaning forward at end expiration. This maneuver brings the aortic root closer to the anterior chest wall. Radiation of the murmur may provide a clue to the cause of the AR. With primary valve disease, such as that due to congenital bicuspid disease, prolapse, or endocarditis, the diastolic murmur tends to radiate along the left sternal border, where it is often louder than appreciated in the second right interspace. When AR is caused by aortic root disease, the diastolic murmur may radiate along the right sternal border. Diseases of the aortic root cause dilation or distortion of the aortic annulus and failure of leaflet coaptation. Causes include Marfan syndrome with aneurysm formation, anuloaortic ectasia, ankylosing spondylitis, and aortic dissection.

Chronic, severe AR also may produce a lower-pitched mid to late, grade 1 or 2 diastolic murmur at the apex (Austin Flint murmur), which is thought to reflect turbulence at the mitral inflow area from the admixture of regurgitant (aortic) and forward (mitral) blood flow. This lower-pitched, apical diastolic murmur can be distinguished from that due to MS by the absence of an opening snap and the response of the murmur to a vasodilator challenge. Lowering afterload with an agent such as amyl nitrite will decrease the duration and magnitude of the aortic-left ventricular diastolic pressure gradient, and thus, the Austin Flint murmur of severe AR will become shorter and softer. The intensity of the diastolic murmur of MS (Fig. 42-6) may either remain constant or increase with afterload reduction because of the reflex increase in cardiac output and mitral valve flow.

Although AS and AR may coexist, a grade 2 or 3 crescendo-decrescendo midsystolic murmur frequently is heard at the base of the heart in patients with isolated, severe AR and is due to an increased volume and rate of systolic flow. Accurate bedside identification of coexistent AS can be difficult unless the carotid pulse examination is abnormal or the midsystolic murmur is of grade 4 or greater intensity. In the absence of heart failure, chronic severe AR is accompanied by several peripheral signs of significant diastolic runoff, including a wide pulse pressure, a "water-hammer" carotid upstroke (Corrigan's pulse), and Quincke's pulsations of the nail beds. The diastolic murmur of

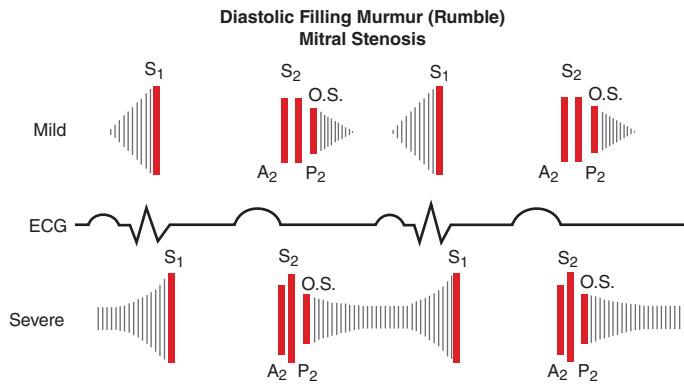


FIGURE 42-6 Diastolic filling murmur (rumble) in mitral stenosis. In mild mitral stenosis, the diastolic gradient across the valve is limited to the phases of rapid ventricular filling in early diastole and presystole. The rumble may occur during either or both periods. As the stenotic process becomes severe, a large pressure gradient exists across the valve during the entire diastolic filling period, and the rumble persists throughout diastole. As the left atrial pressure becomes greater, the interval between A₂ (or P₂) and the opening snap (O.S.) shortens. In severe mitral stenosis, secondary pulmonary hypertension develops and results in a loud P₂ and the splitting interval usually narrows. ECG, electrocardiogram. (Reprinted with permission Examination of the Heart, Part IV: Auscultation of the Heart ©American Heart Association, Inc.)

acute, severe AR is notably shorter in duration and lower pitched than the murmur of chronic AR. It can be very difficult to appreciate in the presence of a rapid heart rate. These attributes reflect the abrupt rate of rise of diastolic pressure within the unprepared and noncompliant left ventricle and the correspondingly rapid decline in the aortic-left ventricular diastolic pressure gradient. Left ventricular diastolic pressure may increase sufficiently to result in premature closure of the mitral valve and a soft first heart sound. Peripheral signs of significant diastolic runoff are generally not present.

Pulmonic regurgitation (PR) results in a decrescendo, early to mid-diastolic murmur (*Graham Steell murmur*) that begins after the pulmonic component of S₂ (P₂), is best heard at the second left interspace, and radiates along the left sternal border. The intensity of the murmur may increase with inspiration. PR is most commonly due to dilation of the valve annulus from chronic elevation of the pulmonary artery pressure. Signs of pulmonary hypertension, including a right ventricular lift and a loud, single or narrowly split S₂, are present. These features also help distinguish PR from AR as the cause of a decrescendo diastolic murmur heard along the left sternal border. PR in the absence of pulmonary hypertension can occur with endocarditis or a congenitally deformed valve. It is usually present after repair of tetralogy of Fallot in childhood. When pulmonary hypertension is not present, the diastolic murmur is softer and lower pitched than the classic Graham Steell murmur, and the severity of the PR can be difficult to appreciate.

TTE is indicated for the further evaluation of a patient with an early to mid-diastolic murmur. Longitudinal assessment of lesion severity, ventricular size, and systolic function helps guide a potential decision for surgical management. TTE also can provide anatomic information regarding the root and proximal ascending aorta, although computed tomographic or magnetic resonance angiography may be indicated for more precise characterization (Chap. 241).

Mid-Diastolic Murmurs (Figs. 42-1F and 42-1G) Mid-diastolic murmurs result from obstruction and/or augmented flow at the level of the mitral or tricuspid valve. Rheumatic fever is the most common cause of MS (Fig. 42-6). In younger patients with pliable valves, S₁ is loud and the murmur begins after an opening snap, which is a high-pitched sound that occurs shortly after S₂. The interval between the pulmonic component of the second heart sound (P₂) and the opening snap is inversely related to the magnitude of the left atrial-left ventricular pressure gradient. The murmur of MS is low-pitched and thus is best heard with the bell of the stethoscope. It is loudest at the left ventricular apex and often is appreciated only when the patient is turned in the left lateral decubitus position. It is usually of grade 1 or 2 intensity

but may be absent when the cardiac output is severely reduced despite significant obstruction. The intensity of the murmur increases during maneuvers that increase cardiac output and mitral valve flow, such as exercise. The duration of the murmur reflects the length of time over which left atrial pressure exceeds left ventricular diastolic pressure. An increase in the intensity of the murmur just before S_1 , a phenomenon known as *presystolic accentuation* (Figs. 42-1A and 42-6), occurs in patients in sinus rhythm and is due to a late increase in transmural flow with atrial contraction. Presystolic accentuation does not occur in patients with atrial fibrillation.

The mid-diastolic murmur associated with tricuspid stenosis is best heard at the lower left sternal border and increases in intensity with inspiration. A prolonged γ descent may be visible in the jugular venous waveform. This murmur is very difficult to hear and most often is obscured by left-sided acoustical events.

There are several other causes of mid-diastolic murmurs. Large left atrial myxomas may prolapse across the mitral valve and cause variable degrees of obstruction to left ventricular inflow (Chap. 271). The murmur associated with an atrial myxoma may change in duration and intensity with changes in body position. An opening snap is not present, and there is no presystolic accentuation. Augmented mitral diastolic flow can occur with isolated severe MR or with a large left-to-right shunt at the ventricular or great vessel level and produce a soft, rapid filling sound (S_3) followed by a short, low-pitched mid-diastolic apical murmur (Fig. 42-1G). The Austin Flint murmur of severe, chronic AR has already been described.

A short, mid-diastolic murmur is rarely heard during an episode of acute rheumatic fever (Carey-Coombs murmur) and probably is due to flow through an edematous mitral valve. An opening snap is not present in the acute phase, and the murmur dissipates with resolution of the acute attack. Complete heart block with dyssynchronous atrial and ventricular activation may be associated with intermittent mid- to late diastolic murmurs if atrial contraction occurs when the mitral valve is partially closed. Mid-diastolic murmurs indicative of increased tricuspid valve flow can occur with severe, isolated TR and with large ASDs and significant left-to-right shunting. Other signs of an ASD are present (Chap. 269), including fixed splitting of S_2 and a midsystolic murmur at the mid- to upper left sternal border. TTE is indicated for evaluation of a patient with a mid- to late diastolic murmur. Findings specific to the diseases discussed above will help guide management.

■ CONTINUOUS MURMURS

(Figs. 42-1H and 42-7) Continuous murmurs begin in systole, peak near the second heart sound, and continue into all or part of diastole. Their presence throughout the cardiac cycle implies a pressure gradient

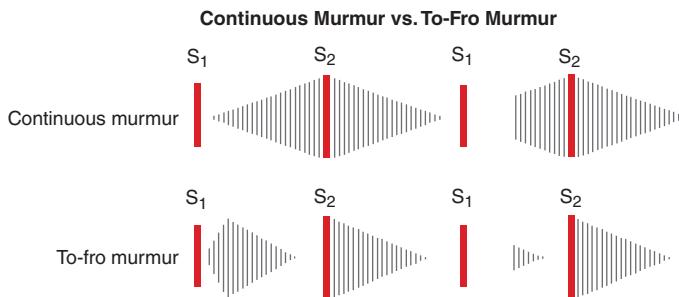


FIGURE 42-7 Comparison of the continuous murmur and the to-fro murmur. During abnormal communication between high-pressure and low-pressure systems, a large pressure gradient exists throughout the cardiac cycle, producing a continuous murmur. A classic example is patent ductus arteriosus. At times, this type of murmur can be confused with a to-fro murmur, which is a combination of systolic ejection murmur and a murmur of semilunar valve incompetence. A classic example of a to-fro murmur is aortic stenosis and regurgitation. A continuous murmur crescendos to near the second heart sound (S_2), whereas a to-fro murmur has two components. The midsystolic ejection component decrescends and disappears as it approaches S_2 . (Reprinted with permission Examination of the Heart, Part IV: Auscultation of the Heart ©American Heart Association, Inc.)

between two chambers or vessels during both systole and diastole. The continuous murmur associated with a patent ductus arteriosus is best heard lateral to the upper left sternal border. Large, uncorrected shunts may lead to pulmonary hypertension, attenuation or obliteration of the diastolic component of the murmur, reversal of shunt flow, and differential cyanosis of the lower extremities. A ruptured sinus of Valsalva aneurysm creates a continuous murmur of abrupt onset at the upper right sternal border. Rupture typically occurs into a right heart chamber, and the murmur is indicative of a continuous pressure difference between the aorta and either the right atrium or the right ventricle. A continuous murmur also may be audible along the left sternal border with a coronary arteriovenous fistula and at the site of an arteriovenous fistula used for hemodialysis access. Enhanced flow through enlarged intercostal collateral arteries in patients with aortic coarctation may produce a continuous murmur along the course of one or more ribs. A cervical bruit with both systolic and diastolic components (a to-fro murmur, Fig. 42-7) usually indicates a high-grade carotid artery stenosis.

Not all continuous murmurs are pathologic. A continuous venous hum can be heard in healthy children and young adults, especially during pregnancy; it is best appreciated in the right supraclavicular fossa and can be obliterated by pressure over the right internal jugular vein or by having the patient turn his or her head toward the examiner. The continuous mammary souffle of pregnancy is created by enhanced arterial flow through engorged breasts and usually appears during the late third trimester or early puerperium. The murmur is louder in systole. Firm pressure with the diaphragm of the stethoscope can eliminate the diastolic portion of the murmur.

■ DYNAMIC AUSCULTATION

(Table 42-2; see Table 239-1) Careful attention to the behavior of heart murmurs during simple maneuvers that alter cardiac hemodynamics can provide important clues to their cause and significance.

Respiration Auscultation should be performed during quiet respiration or with a modest increase in inspiratory effort, as more forceful movement of the chest tends to obscure the heart sounds. Left-sided murmurs may be best heard at end expiration, when lung volumes are minimized, and the heart and great vessels are brought closer to the chest wall. This phenomenon is characteristic of the murmur of AR. Murmurs of right-sided origin, such as tricuspid or pulmonic regurgitation, increase in intensity during inspiration. The intensity of left-sided murmurs either remains constant or decreases with inspiration.

Bedside assessment also should evaluate the behavior of S_2 with respiration and the dynamic relationship between the aortic and pulmonic components (Fig. 42-8). Reversed splitting can be a feature of severe AS, HOCM, left bundle branch block, right ventricular pacing, or acute myocardial ischemia. Fixed splitting of S_2 in the presence of a grade 2 or 3 midsystolic murmur at the mid- or upper left sternal border indicates an ASD. Physiologic but wide splitting during the respiratory cycle implies either premature aortic valve closure, as can occur with severe MR, or delayed pulmonic valve closure due to PS or right bundle branch block.

Alterations of Systemic Vascular Resistance Murmurs can change characteristics after maneuvers that alter systemic vascular

TABLE 42-2 Dynamic Auscultation: Bedside Maneuvers That can be Used to Change the Intensity of Cardiac Murmurs (See Text)

1. Respiration
2. Isometric exercise (handgrip)
3. Transient arterial occlusion
4. Pharmacologic manipulation of preload and/or afterload
5. Valsalva maneuver
6. Rapid standing/squatting
7. Passive leg raising
8. Post-premature beat

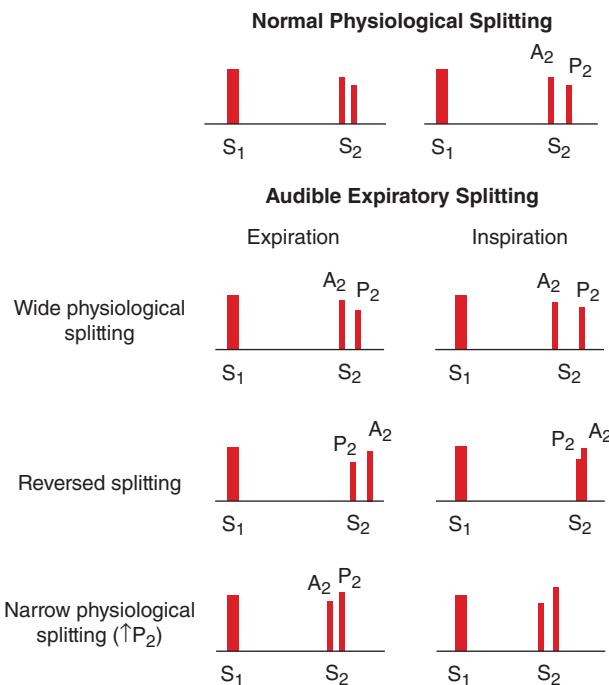


FIGURE 42-8 *Top.* Normal physiologic splitting of the second heart sound. During expiration, the aortic (A_2) and pulmonic (P_2) components of the second heart sound are separated by <30 ms and are appreciated as a single sound. During inspiration, the splitting interval widens, and A_2 and P_2 are clearly separated into two distinct sounds. *Bottom.* Audible expiratory splitting. Wide physiologic splitting is caused by a delay in P_2 (as, for example, with right bundle branch block) or by early closure of the aortic valve (A_2 , as for example with severe mitral regurgitation). Reversed splitting is caused by a delay in A_2 , resulting in paradoxical movement; i.e., with inspiration P_2 moves toward A_2 , and the splitting interval narrows. Narrow physiologic splitting occurs in pulmonary hypertension, and both A_2 and P_2 are heard during expiration at a narrow splitting interval because of the increased intensity and high-frequency composition of P_2 . (Reprinted with permission Examination of the Heart, Part IV: Auscultation of the Heart ©American Heart Association, Inc.)

resistance and left ventricular afterload. The systolic murmurs of MR and VSD become louder during sustained handgrip, simultaneous inflation of blood pressure cuffs on both upper extremities to pressures 20–40 mmHg above systolic pressure for 20 s, or infusion of a vasopressor agent. The murmurs associated with AS or HOCM will become softer or remain unchanged with these maneuvers. The diastolic murmur of AR becomes louder in response to interventions that raise systemic vascular resistance.

Opposite changes in systolic and diastolic murmurs may occur with the use of pharmacologic agents that lower systemic vascular resistance. Inhaled amyl nitrite is now rarely used for this purpose but can help distinguish the murmur of AS or HOCM from that of either MR or VSD, if necessary. The former two murmurs increase in intensity, whereas the latter two become softer after exposure to amyl nitrite. As noted previously, the Austin Flint murmur of severe AR becomes softer, but the mid-diastolic rumble of MS becomes louder, in response to the abrupt lowering of systemic vascular resistance with amyl nitrite and enhanced transmural valve flow.

Changes in Venous Return The Valsalva maneuver results in an increase in intrathoracic pressure, followed by a decrease in venous return, ventricular filling, and cardiac output. The majority of murmurs decrease in intensity during the strain phase of the maneuver. Two notable exceptions are the murmurs associated with MVP and HOCM, both of which become louder during the Valsalva maneuver. The murmur of MVP may also become longer as leaflet prolapse occurs earlier in systole at smaller ventricular volumes. These murmurs behave in a similar and parallel fashion with standing. Both the click and the murmur of MVP move closer in timing to S_1 on rapid standing from a squatting position (Fig. 42-3). The increase in the

intensity of the murmur of HOCM is predicated on the augmentation of the dynamic left ventricular outflow tract gradient that occurs with reduced ventricular filling. Squatting results in abrupt increases in both venous return (preload) and left ventricular afterload that increase ventricular volume, changes that predictably cause a decrease in the intensity and duration of the murmurs associated with MVP and HOCM; the click and murmur of MVP move away from S_1 with squatting. Passive leg raising can be used to increase venous return in patients who are unable to squat and stand. This maneuver may lead to a decrease in the intensity of the murmur associated with HOCM but has less effect in patients with MVP.

Post-Premature Ventricular Contraction A change in the intensity of a systolic murmur in the first beat after a premature beat, or in the beat after a long cycle length in patients with atrial fibrillation, can help distinguish AS from MR, particularly in an older patient in whom the murmur of AS is well transmitted to the apex. Systolic murmurs due to left ventricular outflow obstruction, including that due to AS, increase in intensity in the beat after a premature beat because of the combined effects of enhanced left ventricular filling and post-extrasystolic potentiation of contractile function. Forward flow accelerates, causing an increase in the gradient and a louder murmur. The intensity of the murmur of MR does not change in the post-premature beat as there is relatively little further increase in mitral valve flow or change in the left ventricular-left atrial gradient.

THE CLINICAL CONTEXT

Additional clues to the etiology and importance of a heart murmur can be gleaned from the history and other physical examination findings. Symptoms suggestive of cardiovascular, neurologic, or pulmonary disease help focus the differential diagnosis, as do findings relevant to the jugular venous pressure and waveforms, the arterial pulses, other heart sounds, the lungs, the abdomen, the skin, and the extremities. In many instances, laboratory studies, an ECG, and/or a chest x-ray may have been obtained earlier and may contain valuable information. A patient with suspected infective endocarditis, for example, may have a murmur in the setting of fever, chills, anorexia, fatigue, dyspnea, splenomegaly, petechiae, and positive blood cultures. A new systolic murmur in a patient with a marked fall in blood pressure after a recent MI suggests myocardial rupture. By contrast, an isolated grade 1 or 2 midsystolic murmur at the left sternal border in a healthy, active, and asymptomatic young adult is most likely a benign finding for which no further evaluation is indicated. The context in which the murmur is appreciated often dictates the need for further testing and the pace of the evaluation.

ECHOCARDIOGRAPHY

(**Fig. 42-9; Chaps. 239 and 241**) Echocardiography with color flow and spectral Doppler is a valuable tool for the assessment of cardiac murmurs. Information regarding valve structure and function, chamber size, wall thickness, ventricular function, estimated pulmonary artery pressures, intracardiac shunt flow, pulmonary and hepatic vein flow, and aortic flow can be ascertained readily. It is important to note that Doppler signals of trace or mild valvular regurgitation of no clinical consequence can be detected with structurally normal tricuspid, pulmonic, and mitral valves. Such signals are not likely to generate enough turbulence to create an audible murmur.

Echocardiography is indicated for the evaluation of patients with early, late, or holosystolic murmurs and patients with grade 3 or louder midsystolic murmurs. Patients with grade 1 or 2 midsystolic murmurs but other symptoms or signs of cardiovascular disease, including those from ECG or chest x-ray, should also undergo echocardiography. Echocardiography is also indicated for the evaluation of any patient with a diastolic murmur and for patients with continuous murmurs not due to a venous hum or mammary souffle. Echocardiography should be considered when there is a clinical need to verify normal cardiac structure and function in a patient whose symptoms and signs are probably noncardiac in origin. The performance of serial

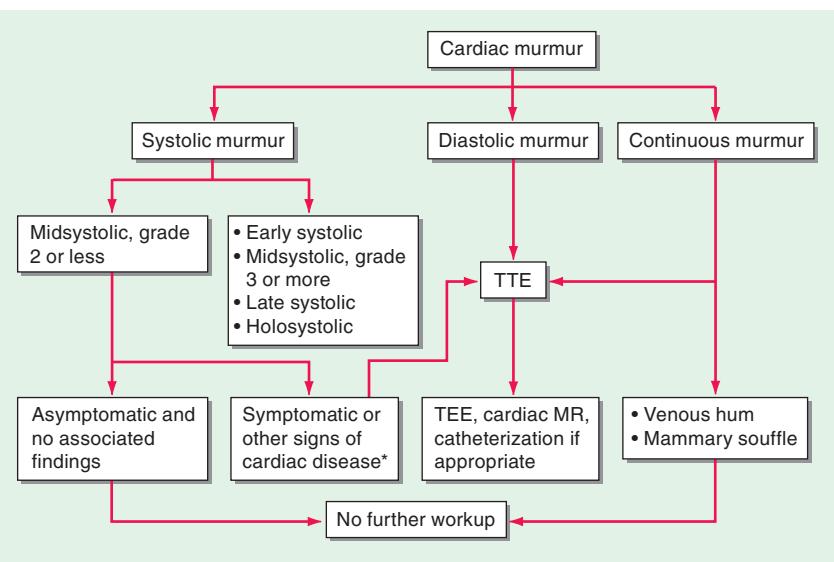


FIGURE 42-9 Strategy for evaluating heart murmurs. *If an electrocardiogram or chest x-ray has been obtained and is abnormal, echocardiography is indicated. MR, magnetic resonance; TEE, transesophageal echocardiography; TTE, transthoracic echocardiography. (Adapted from RO Bonow et al: 1998 ACC/AHA Guideline for the management of patients with valvular heart disease. *J Am Coll Cardiol* 32:1486, 1998.)

echocardiography to follow the course of asymptomatic individuals with valvular heart disease is a central feature of their longitudinal assessment, and it provides valuable information that may have an impact on decisions regarding the timing of surgery. Routine echocardiography is *not* recommended for asymptomatic patients with a grade 1 or 2 midsystolic murmur without other signs of heart disease. For this category of patients, referral to a cardiovascular specialist could be considered if there is doubt about the significance of the murmur after the initial examination.

The selective use of echocardiography outlined above has not been subjected to rigorous analysis of its cost-effectiveness. For some clinicians, handheld or miniaturized cardiac ultrasound devices have replaced the stethoscope. Although several reports attest to the improved sensitivity of such devices for the detection of valvular heart disease (e.g., rheumatic heart disease in susceptible populations), accuracy is highly operator-dependent, and incremental cost considerations and outcomes have not been addressed adequately for most patient scenarios. The use of electronic or digital stethoscopes with spectral display capabilities has also been proposed as a method to improve the characterization of heart murmurs and the mentored teaching of cardiac auscultation.

■ OTHER CARDIAC TESTING

(**Chap. 241**, Fig. 42-9) In relatively few patients, clinical assessment and TTE do not adequately characterize the origin and significance of a heart murmur. Transesophageal echocardiography (TEE) can be considered for further evaluation, especially when the TTE windows are limited by body size, chest configuration, or intrathoracic pathology. TEE offers enhanced sensitivity for the detection of a wide range of structural cardiac disorders. Electrocardiographically gated cardiac magnetic resonance (CMR) imaging can provide quantitative information regarding valvular function, regurgitant fraction, regurgitant volume, shunt flow, chamber and great vessel size, ventricular function, and myocardial perfusion. CMR imaging has largely supplanted the need for cardiac catheterization and invasive hemodynamic assessment when there is a discrepancy between the clinical and echocardiographic findings in patients with regurgitant heart valve disease, such as MR or AR. Both CMR and cardiac CT can provide assessment of aortic valve leaflet number when there is uncertainty by TTE regarding whether the valve is bi- or tricuspid, as well as provide information on aortic root and ascending aortic anatomy. The use of coronary CT angiography to exclude coronary artery disease in selected patients with a low pretest probability of disease before valve surgery has gained

wider acceptance. Invasive angiography and hemodynamic assessment may be required for a more complete preoperative evaluation.

■ INTEGRATED APPROACH

The accurate identification of a heart murmur begins with a systematic approach to cardiac auscultation. Characterization of its major attributes, as reviewed above, allows the examiner to construct a preliminary differential diagnosis, which is then refined by integration of information available from the history, associated cardiac findings, the general physical examination, and the clinical context. The need for and urgency of further testing follow sequentially. Correlation of the findings on auscultation with the noninvasive data provides an educational feedback loop and an opportunity for improving physical examination skills. Cost considerations mandate that noninvasive imaging be justified on the basis of its incremental contribution to diagnosis, treatment, and outcome. Cardiac auscultation using a stethoscope remains a time-honored tradition in medicine, the benefits of which extend beyond accurate recognition of heart sounds. Selective augmentation with, rather than wholesale replacement by, handheld ultrasound and newer technologies may improve diagnostic accuracy and better guide therapeutic decisions.

■ FURTHER READING

- EDELMAN ER, WEBER BN: Tenuous tether. *N Engl J Med* 373:2199, 2015.
- EVANGELISTA A et al: Hand-held cardiac ultrasound screening performed by family doctors with remote expert support interpretation. *Heart* 102:376, 2016.
- FANG LC, O'GARA PT: The history and physical examination. An evidence-based approach, in *Braunwald's Heart Disease. A Textbook of Cardiovascular Medicine*, 11th ed, DP Zipes et al (eds). Philadelphia, Elsevier/Saunders, 2019, pp 83-101.
- FUSTER V: The stethoscope's prognosis. Very much alive and very necessary. *J Am Coll Cardiol* 67:1118, 2016.
- OTTO CM et al: 2020 AHA/ACC guideline for the management of patients with valvular heart disease. *J Am Coll Cardiol* 143:e72, 2021.
- STOKKE TM et al: Brief group training of medical students in focused cardiac ultrasound may improve diagnostic accuracy of physical examination. *J Am Soc Echocardiogr* 27:1238, 2014.



Palpitations are extremely common among patients who present to their internists and can best be defined as a “thumping,” “pounding,” or “fluttering” sensation in the chest. This sensation can be either intermittent or sustained and either regular or irregular. Most patients interpret palpitations as an unusual awareness of the heartbeat and become especially concerned when they sense that they have had “skipped” or “missing” heartbeats. Palpitations are often noted when the patient is quietly resting, during which time other stimuli are minimal. Palpitations that are positional generally reflect a structural process within (e.g., atrial myxoma) or adjacent to (e.g., mediastinal mass) the heart.

Palpitations are brought about by cardiac (43%), psychiatric (31%), miscellaneous (10%), and unknown (16%) causes, according to one large series. Among the cardiovascular causes are premature atrial and ventricular contractions, supraventricular and ventricular arrhythmias, mitral valve prolapse (with or without associated arrhythmias), aortic insufficiency, atrial myxoma, myocarditis, and pulmonary embolism. Intermittent palpitations are commonly caused by premature atrial or ventricular contractions: the post-extrasystolic beat is sensed by the patient owing to the increase in ventricular end-diastolic dimension following the pause in the cardiac cycle and the increased strength of contraction (post-extrasystolic potentiation) of that beat. Regular, sustained palpitations can be caused by regular supraventricular and ventricular tachycardias. Irregular, sustained palpitations can be caused by atrial fibrillation. It is important to note that most arrhythmias are not associated with palpitations. In those that are, it is often useful either to ask the patient to “tap out” the rhythm of the palpitations or to take his or her pulse during palpitations. In general, hyperdynamic cardiovascular states caused by catecholaminergic stimulation from exercise, stress, or pheochromocytoma can lead to palpitations. Palpitations are common among athletes, especially older endurance athletes. In addition, the enlarged ventricle of aortic regurgitation and accompanying hyperdynamic precordium frequently lead to the sensation of palpitations. Other factors that enhance the strength of myocardial contraction, including tobacco, caffeine, aminophylline, atropine, thyroxine, cocaine, and amphetamines, can cause palpitations.

Psychiatric causes of palpitations include panic attacks or disorders, anxiety states, and somatization, alone or in combination. Patients with psychiatric causes for palpitations more commonly report a longer duration of the sensation (>15 min) and other accompanying symptoms than do patients with other causes. Among the miscellaneous causes of palpitations are thyrotoxicosis, drugs (see above) and ethanol, spontaneous skeletal muscle contractions of the chest wall, pheochromocytoma, and systemic mastocytosis.

APPROACH TO THE PATIENT

Palpitations

The principal goal in assessing patients with palpitations is to determine whether the symptom is caused by a life-threatening arrhythmia. Patients with preexisting coronary artery disease (CAD) or risk factors for CAD are at greatest risk for ventricular arrhythmias (**Chap. 246**) as a cause for palpitations. In addition, the association of palpitations with other symptoms suggesting hemodynamic compromise, including syncope or lightheadedness, supports this diagnosis. Palpitations caused by sustained tachyarrhythmias in patients with CAD can be accompanied by angina pectoris or dyspnea, and, in patients with ventricular dysfunction (systolic or diastolic), aortic stenosis, hypertrophic cardiomyopathy, or mitral stenosis (with or without CAD), can be accompanied by dyspnea from increased left atrial and pulmonary venous pressure.

Key features of the physical examination that will help confirm or refute the presence of an arrhythmia as a cause for palpitations (as well as its adverse hemodynamic consequences) include measurement of the vital signs, assessment of the jugular venous pressure and pulse, and auscultation of the chest and precordium. A resting electrocardiogram can be used to document the arrhythmia. If exertion is known to induce the arrhythmia and accompanying palpitations, exercise electrocardiography can be used to make the diagnosis. If the arrhythmia is sufficiently infrequent, other methods must be used, including continuous electrocardiographic (Holter) monitoring; telephonic monitoring, through which the patient can transmit an electrocardiographic tracing during a sensed episode; loop recordings (external or implantable), which can capture the electrocardiographic event for later review; and mobile (self-monitoring) cardiac outpatient telemetry. Data suggest that Holter monitoring is of limited clinical utility, while the implantable loop recorder and mobile cardiac outpatient telemetry are safe and

possibly more cost-effective in the assessment of patients with (infrequent) recurrent, unexplained palpitations. The use of a diary or an electronic marker to indicate the timing of palpitations sensed by the patient is essential for appropriate interpretation of these studies.

Most patients with palpitations do not have serious arrhythmias or underlying structural heart disease. If sufficiently troubling to the patient, occasional benign atrial or ventricular premature contractions can often be managed with beta-blocker therapy. Palpitations incited by alcohol, tobacco, or illicit drugs need to be managed by abstention, while those caused by pharmacologic agents should be addressed by considering alternative therapies when appropriate or possible. Psychiatric causes of palpitations may benefit from cognitive therapy or pharmacotherapy. The physician should note that palpitations are at the very least bothersome and, on occasion, frightening to the patient. Once serious causes for the symptom have been excluded, the patient should be reassured that the palpitations will not adversely affect prognosis.

FURTHER READING

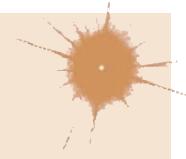
- CROSSLAND S, BERKIN L: Problem based review: The patient with palpitations. *Acute Med* 11:169, 2012.
- JAMSHED N et al: Emergency management of palpitations in the elderly: Epidemiology, diagnostic approaches, and therapeutic options. *Clin Geriatr Med* 29:205, 2013.
- MARTSON HR et al: Mobile self-monitoring ECG devices to diagnose arrhythmias that coincide with palpitations: A scoping review. *Healthcare (Basel)* 7:pii: E96, 2019.
- SAKH R et al: Insertable cardiac monitors: current indications and devices. *Expert Rev Med Devices* 16:45, 2019.
- WEBER BE, KAPOOR WN: Evaluation and outcomes of patients with palpitations. *Am J Med* 100:138, 1996.

Section 6 Alterations in Gastrointestinal Function

44

Dysphagia

Ikuo Hirano, Peter J. Kahrilas



Dysphagia—difficulty with swallowing—refers to problems with the transit of food or liquid from the mouth to the hypopharynx or through the esophagus. Severe dysphagia can compromise nutrition, cause aspiration, and reduce quality of life. Additional terminology pertaining to swallowing dysfunction is as follows. *Aphagia* (inability to swallow) typically denotes complete esophageal obstruction, most commonly encountered in the acute setting of a food bolus or foreign body impaction. *Odynophagia* refers to painful swallowing, typically resulting from mucosal ulceration within the oropharynx or esophagus. It commonly is accompanied by dysphagia, but the converse is not true. *Globus pharyngeus* is a foreign body sensation localized in the neck that does not interfere with swallowing and sometimes is relieved by swallowing. *Transfer dysphagia* frequently results in nasal regurgitation or pulmonary aspiration during swallowing and is characteristic of oropharyngeal dysphagia. *Phagophobia* (fear of swallowing) and *refusal to swallow* may be psychogenic or related to anticipatory anxiety about food bolus obstruction, odynophagia, or aspiration.

PHYSIOLOGY OF SWALLOWING

Swallowing begins with a voluntary (oral) phase that includes preparation during which food is masticated and mixed with saliva. This is followed by a transfer phase during which the bolus is pushed into the

pharynx by the tongue. Bolus entry into the hypopharynx initiates the pharyngeal swallow response, which is centrally mediated and involves a complex series of actions, the net result of which is to propel food through the pharynx into the esophagus while preventing its entry into the airway. To accomplish this, the larynx is elevated and pulled forward, actions that also facilitate upper esophageal sphincter (UES) opening. Tongue pulsion then propels the bolus through the UES, followed by a peristaltic contraction that clears residue from the pharynx and through the esophagus. The lower esophageal sphincter (LES) relaxes as the food enters the esophagus and remains relaxed until the peristaltic contraction has delivered the bolus into the stomach. Peristaltic contractions elicited in response to a swallow are called *primary peristalsis* and involve sequenced inhibition followed by contraction of the musculature along the entire length of the esophagus. The inhibition that precedes the peristaltic contraction is called *deglutitive inhibition*. Local distention of the esophagus anywhere along its length, as may occur with gastroesophageal reflux, activates *secondary peristalsis* that begins at the point of distention and proceeds distally. Tertiary esophageal contractions are nonperistaltic, disordered esophageal contractions that may be observed to occur spontaneously during fluoroscopic observation.

The musculature of the oral cavity, pharynx, UES, and cervical esophagus is striated and directly innervated by lower motor neurons carried in cranial nerves (Fig. 44-1). Oral cavity muscles are innervated by the fifth (trigeminal) and seventh (facial) cranial nerves; the tongue, by the twelfth (hypoglossal) cranial nerve. Pharyngeal muscles are innervated by the ninth (glossopharyngeal) and tenth (vagus) cranial nerves.

Physiologically, the UES consists of the cricopharyngeus muscle, the adjacent inferior pharyngeal constrictor, and the proximal portion of the cervical esophagus. UES innervation is derived from the vagus nerve, whereas the innervation to the musculature acting on the UES to facilitate its opening during swallowing comes from the fifth, seventh, and twelfth cranial nerves. The UES remains closed at rest owing to both its inherent elastic properties and neurogenically mediated contraction of the cricopharyngeus muscle. UES opening during swallowing involves both cessation of vagal excitation to the

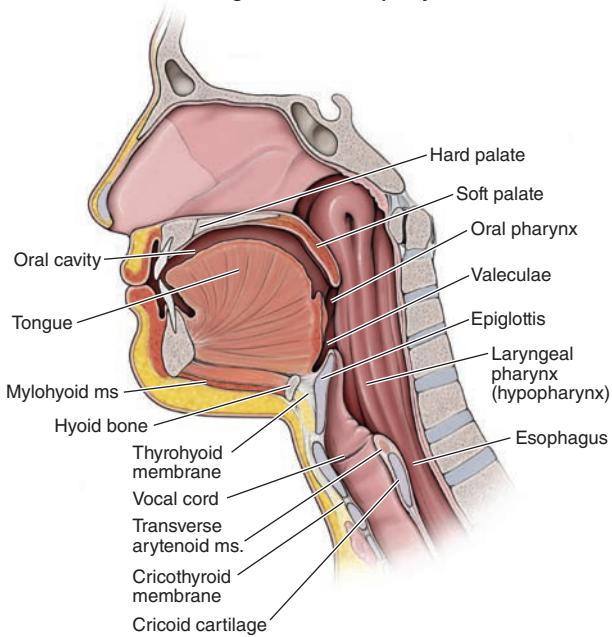
cricopharyngeus and simultaneous contraction of the suprathyroid and geniohyoid muscles that pull open the UES in conjunction with the upward and forward displacement of the larynx.

The neuromuscular apparatus for peristalsis is distinct in proximal and distal parts of the esophagus. The cervical esophagus, like the pharyngeal musculature, consists of striated muscle and is directly innervated by lower motor neurons of the vagus nerve. Peristalsis in the proximal esophagus is governed by the sequential activation of the vagal motor neurons in the nucleus ambiguus. In contrast, the distal esophagus and LES are composed of smooth muscle and are controlled by excitatory and inhibitory neurons within the esophageal myenteric plexus. Medullary preganglionic neurons from the dorsal motor nucleus of the vagus trigger peristalsis via these ganglionic neurons during primary peristalsis. Neurotransmitters of the excitatory ganglionic neurons are acetylcholine and substance P; those of the inhibitory neurons are vasoactive intestinal peptide and nitric oxide. Peristalsis results from the patterned activation of inhibitory followed by excitatory ganglionic neurons, with progressive dominance of the inhibitory neurons distally. Similarly, LES relaxation occurs with the onset of deglutitive inhibition and persists until the peristaltic sequence is complete. At rest, the LES is contracted because of excitatory ganglionic stimulation and its intrinsic myogenic tone, a property that distinguishes it from the adjacent esophagus. The function of the LES is supplemented by the surrounding muscle of the right diaphragmatic crus, which acts as an external sphincter during inspiration, cough, or abdominal straining.

PATHOPHYSIOLOGY OF DYSPHAGIA

Dysphagia can be subclassified both by location and by the circumstances in which it occurs. With respect to location, distinct considerations apply to oral, pharyngeal, or esophageal dysphagia. Normal transport of an ingested bolus depends on the consistency and size of the bolus, the caliber of the lumen, the integrity of peristaltic contraction, and deglutitive inhibition of both the UES and the LES. Dysphagia caused by an oversized bolus or a narrow lumen is called *structural dysphagia*, whereas dysphagia due to abnormalities of peristalsis or impaired sphincter relaxation after swallowing is called *propulsive* or

Sagittal view of the pharynx



Musculature of the pharynx

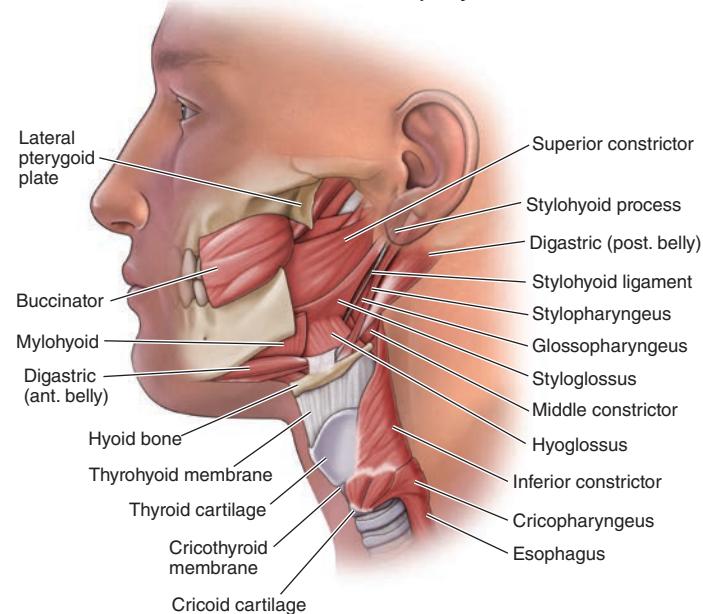


FIGURE 44-1 Sagittal and diagrammatic views of the musculature involved in enacting oropharyngeal swallowing. Note the dominance of the tongue in the sagittal view and the intimate relationship between the entrance to the larynx (airway) and the esophagus. In the resting configuration illustrated, the esophageal inlet is closed. This is transiently reconfigured such that the esophageal inlet is open and the laryngeal inlet closed during swallowing. (Adapted from PJ Kahrilas, in DW Gelfand and JE Richter [eds]: *Dysphagia: Diagnosis and Treatment*. New York, Igaku-Shoin Medical Publishers, 1989, pp. 11–28.)

motor dysphagia. More than one mechanism may be operative in a patient with dysphagia. Scleroderma commonly presents with absent peristalsis as well as a weakened LES that predisposes patients to peptic stricture formation. Likewise, radiation therapy for head and neck cancer may compound the functional deficits in the oropharyngeal swallow attributable to the tumor and cause cervical esophageal stenosis. It is worth noting that in addition to bolus transit, symptom reporting of dysphagia is dependent upon intact sensory innervation and central nervous system perception.

Oral and Pharyngeal (Oropharyngeal) Dysphagia Oral-phase dysphagia is associated with poor bolus formation and control so that food has prolonged retention within the oral cavity and may seep out of the mouth. Drooling and difficulty in initiating swallowing are other characteristic signs. Poor bolus control also may lead to premature spillage of food into the hypopharynx with resultant aspiration into the trachea or regurgitation into the nasal cavity. Pharyngeal-phase dysphagia is associated with retention of food in the pharynx due to poor tongue or pharyngeal propulsion or obstruction at the UES. Signs and symptoms of concomitant hoarseness or cranial nerve dysfunction may be associated with oropharyngeal dysphagia.

Oropharyngeal dysphagia may be due to neurologic, muscular, structural, iatrogenic, infectious, and metabolic causes. Iatrogenic, neurologic, and structural pathologies are most common. Iatrogenic causes include surgery and radiation, often in the setting of head and neck cancer. Neurogenic dysphagia resulting from cerebrovascular accidents, Parkinson's disease, and amyotrophic lateral sclerosis is a major source of morbidity related to aspiration and malnutrition. Medullary nuclei directly innervate the oropharynx. Lateralization of pharyngeal dysphagia implies either a structural pharyngeal lesion or a neurologic process that selectively targeted the ipsilateral brainstem nuclei or cranial nerve. Advances in functional brain imaging have elucidated an important role of the cerebral cortex in swallow function and dysphagia. Asymmetry in the cortical representation of the pharynx provides an explanation for the dysphagia that occurs as a consequence of unilateral cortical cerebrovascular accidents.

Oropharyngeal structural lesions causing dysphagia include Zenker's diverticulum, cricopharyngeal bar, and neoplasia. Zenker's diverticulum typically is encountered in elderly patients. In addition to dysphagia, patients may present with regurgitation of particulate food debris, aspiration, and halitosis. The pathogenesis is related to stenosis of the cricopharyngeus that causes diminished opening of the UES and results in increased hypopharyngeal pressure during swallowing with development of a pulsion diverticulum immediately above the cricopharyngeus in a region of potential weakness known as Killian's dehiscence. A cricopharyngeal bar, appearing as a prominent indentation behind the lower third of the cricoid cartilage, is related to Zenker's diverticulum in that it involves limited distensibility of the cricopharyngeus and can lead to the formation of a Zenker's diverticulum. However, a cricopharyngeal bar is a common radiographic finding, and most patients with transient cricopharyngeal bars are asymptomatic, making it important to rule out alternative etiologies of dysphagia before treatment. Furthermore, cricopharyngeal bars may be secondary to other neuromuscular disorders that impair opening of the UES.

Since the pharyngeal phase of swallowing occurs in less than a second, rapid-sequence fluoroscopy is necessary to evaluate for functional abnormalities. Adequate fluoroscopic examination requires that the patient be conscious and cooperative. The study incorporates recordings of swallow sequences during ingestion of food and liquids of varying consistencies. The pharynx is examined to detect bolus retention, regurgitation into the nose, or aspiration into the trachea. Timing and integrity of pharyngeal contraction and opening of the UES with a swallow are analyzed to assess both aspiration risk and the potential for swallow therapy. Structural abnormalities of the oropharynx, especially those that may require biopsies, also should be assessed by direct laryngoscopic examination.

Esophageal Dysphagia The adult esophagus measures 18–26 cm in length and is anatomically divided into the cervical esophagus, extending from the pharyngoesophageal junction to the suprasternal notch, and the thoracic esophagus, which continues to the diaphragmatic hiatus. When distended, the esophageal lumen has internal dimensions of about 2 cm in the anteroposterior plane and 3 cm in the lateral plane. Solid food dysphagia becomes common when the lumen is narrowed to <13 mm, but also can occur with larger diameters in the setting of poorly masticated food or motor dysfunction. Circumferential lesions are more likely to cause dysphagia than are lesions that involve only a partial circumference of the esophageal wall. The most common structural causes of dysphagia are Schatzki's rings, eosinophilic esophagitis, and peptic strictures. Dysphagia also occurs in the setting of gastroesophageal reflux disease without a stricture, perhaps on the basis of altered esophageal sensation, reduced esophageal mural distensibility, or motor dysfunction.

Propulsive disorders leading to esophageal dysphagia result from abnormalities of peristalsis and/or degluttitive inhibition, potentially affecting the cervical or thoracic esophagus. Since striated muscle pathology usually involves both the oropharynx and the cervical esophagus, the clinical manifestations usually are dominated by oropharyngeal dysphagia. Diseases affecting smooth muscle involve both the thoracic esophagus and the LES. A dominant manifestation of this, absent peristalsis, refers to either the complete absence of swallow-induced contraction (absent contractility) or the presence of nonperistaltic, disordered contractions. Absent peristalsis and failure of degluttitive LES relaxation are the defining features of achalasia. In diffuse esophageal spasm (DES), LES function is normal, with the disordered motility restricted to the esophageal body. Absent contractility combined with severe weakness of the LES is a pattern commonly found in patients with scleroderma.

APPROACH TO THE PATIENT

Dysphagia

Figure 44-2 shows an algorithm for the approach to a patient with dysphagia.

HISTORY

The patient history is extremely valuable in making a presumptive diagnosis or at least substantially limiting the differential diagnoses in most patients. Key elements of the history are the localization of dysphagia, the circumstances in which dysphagia is experienced, other symptoms associated with dysphagia, and progression. Dysphagia that localizes to the suprasternal notch may indicate either an oropharyngeal or an esophageal etiology as distal dysphagia is referred proximally about 30% of the time. Dysphagia that localizes to the chest is esophageal in origin. Nasal regurgitation and tracheobronchial aspiration manifest by coughing with swallowing are hallmarks of oropharyngeal dysphagia. Severe cough with swallowing may also be a sign of a tracheoesophageal fistula. The presence of hoarseness may be another important diagnostic clue. When hoarseness precedes dysphagia, the primary lesion is usually laryngeal; hoarseness that occurs after the development of dysphagia may result from compromise of the recurrent laryngeal nerve by a malignancy. The type of food causing dysphagia is an important consideration. Intermittent dysphagia that occurs only with solid food implies structural dysphagia, whereas constant dysphagia with both liquids and solids strongly suggests an esophageal motor abnormality. Two caveats to this pattern are that despite having a motor abnormality, patients with scleroderma generally develop mild dysphagia for solids only and that patients with oropharyngeal dysphagia often have greater difficulty managing liquids than solids. Dysphagia that is progressive over the course of weeks to months raises concern for neoplasia. Episodic dysphagia to solids that is unchanged or slowly progressive over years indicates a benign disease process such as a Schatzki ring or eosinophilic

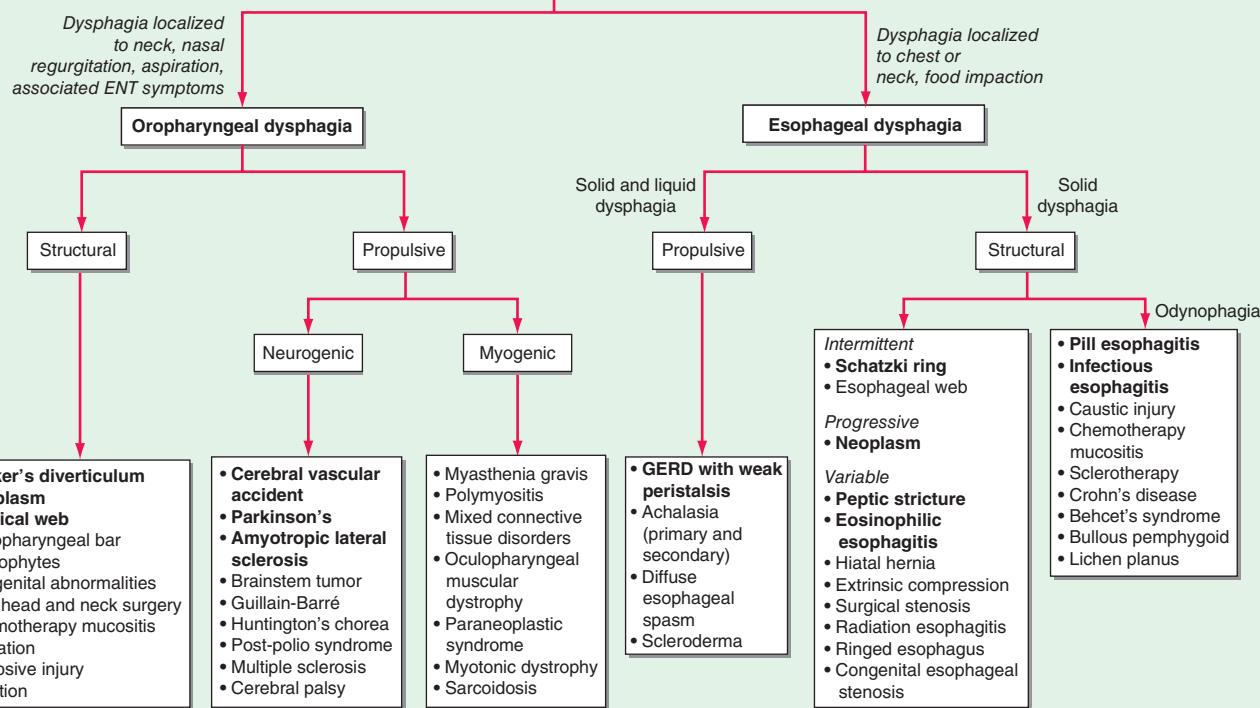


FIGURE 44-2 Approach to the patient with dysphagia. Etiologies in bold print are the most common. ENT, ear, nose, and throat; GERD, gastroesophageal reflux disease.

esophagitis. Food impaction with a prolonged inability to pass an ingested bolus even with ingestion of liquid is typical of a structural dysphagia. Chest pain may accompany dysphagia whether it is related to motor disorders, structural disorders, or reflux disease. A prolonged history of heartburn preceding the onset of dysphagia is suggestive of peptic stricture and, infrequently, esophageal adenocarcinoma. A history of prolonged nasogastric intubation, esophageal or head and neck surgery, ingestion of caustic agents or pills, previous radiation or chemotherapy, or associated mucocutaneous diseases may help isolate the cause of dysphagia. With accompanying odynophagia, which usually is indicative of ulceration, infectious or pill-induced esophagitis should be suspected. In patients with AIDS or other immunocompromised states, esophagitis due to opportunistic infections such as *Candida*, herpes simplex virus, or cytomegalovirus and to tumors such as Kaposi's sarcoma and lymphoma should be considered. A history of atopy increases concerns for eosinophilic esophagitis, which is most prevalent in Caucasian male patients between the ages of 20 and 40 years. Medication use should identify agents associated with pill esophagitis and narcotics that are associated with opioid-induced esophageal dysmotility.

PHYSICAL EXAMINATION

Physical examination is important in the evaluation of oral and pharyngeal dysphagia because dysphagia is usually only one of many manifestations of a more global disease process. Signs of bulbar or pseudobulbar palsy, including dysarthria, dysphonia, ptosis, and tongue atrophy, in addition to evidence of generalized neuromuscular disease, should be elicited. The neck should be examined for thyromegaly or lymphadenopathy. A careful inspection of the mouth and pharynx should disclose inflammatory or infectious lesions. Missing dentition can interfere with mastication and exacerbate an existing cause of dysphagia. Physical examination is less helpful in the evaluation of esophageal dysphagia as most relevant

pathology is restricted to the esophagus. The notable exception is skin disease. Changes in the skin and oral mucosa may suggest a diagnosis of scleroderma or mucocutaneous diseases such as pemphigoid, lichen planus, and epidermolysis bullosa, all of which can involve the esophagus.

DIAGNOSTIC PROCEDURES

Although most instances of dysphagia are attributable to benign disease processes, dysphagia is also a cardinal symptom of several malignancies, making it an important symptom to evaluate. Cancer may result in dysphagia most commonly as the result of intraluminal obstruction (esophageal or proximal gastric cancer, metastatic deposits) and less commonly due to extrinsic compression (lymphoma, lung cancer) or paraneoplastic syndromes. Even when not attributable to malignancy, dysphagia is usually a manifestation of an identifiable and treatable disease entity, making its evaluation beneficial to the patient and gratifying to the practitioner. The specific diagnostic algorithm to pursue is guided by the details of the history (Fig. 44-2). If oral or pharyngeal dysphagia is suspected, a fluoroscopic swallow study, usually done by a swallow therapist, is the procedure of choice. Otolaryngoscopic and neurologic evaluation also can be important, depending on the circumstances. For suspected esophageal dysphagia, upper endoscopy is the single most useful test. Endoscopy allows better visualization of mucosal lesions than does barium radiography and also allows for procurement of mucosal biopsies. Endoscopic or histologic abnormalities are evident in the leading causes of esophageal dysphagia: Schatzki's ring, gastroesophageal reflux disease, and eosinophilic esophagitis. Furthermore, therapeutic intervention with esophageal dilation can be done as part of the procedure if it is deemed necessary. The emergence of eosinophilic esophagitis as a leading cause of dysphagia in both children and adults has led to the recommendation that esophageal mucosal biopsies be obtained routinely in the

evaluation of unexplained dysphagia even if characteristic, endoscopically identified esophageal mucosal features are absent. For cases of suspected esophageal motility disorders, endoscopy is still the appropriate initial evaluation as neoplastic and inflammatory conditions can secondarily produce patterns of either achalasia or esophageal spasm. Esophageal manometry is done if dysphagia is not adequately explained by endoscopy or to confirm the diagnosis of a suspected esophageal motor disorder. Barium radiography can provide useful adjunctive information in cases of subtle or complex esophageal strictures, prior esophageal surgery, esophageal diverticula, or paraesophageal herniation. Use of a barium tablet in conjunction with fluoroscopy can identify strictures and esophageal motility disorders that may be overlooked with liquid barium. In specific cases, computed tomography (CT) examination, esophageal manometry with solid meal challenge, and endoscopic ultrasonography may be useful.

TREATMENT

Treatment of dysphagia depends on both the locus and the specific etiology. Oropharyngeal dysphagia most commonly results from functional deficits caused by neurologic disorders. In such circumstances, the treatment focuses on utilizing postures or maneuvers devised to reduce pharyngeal residue and enhance airway protection learned under the direction of a swallow therapist. Aspiration risk may be reduced by altering the consistency of ingested food and liquid. Dysphagia resulting from a cerebrovascular accident usually, but not always, spontaneously improves within the first few weeks after the event. More severe and persistent cases may require consideration of gastrostomy and enteral feeding. Patients with myasthenia gravis (*Chap. 448*) and polymyositis (*Chap. 365*) may respond to medical treatment of the primary neuromuscular disease. Surgical intervention with cricopharyngeal myotomy is usually not helpful, with the exception of specific disorders such as symptomatic cricopharyngeal bar, Zenker's diverticulum, and oculopharyngeal muscular dystrophy. Chronic neurologic disorders such as Parkinson's disease and amyotrophic lateral sclerosis may manifest with severe oropharyngeal dysphagia. Feeding by a nasogastric tube or an endoscopically placed gastrostomy tube may be considered for nutritional support; however, these maneuvers do not provide protection against aspiration of salivary secretions or refluxed gastric contents.

Treatment of esophageal dysphagia is covered in detail in *Chap. 323*. The majority of causes of structural, esophageal dysphagia are effectively managed by means of esophageal dilation using bougie or balloon dilators. Cancer and achalasia are often managed surgically, although endoscopic techniques are available for both palliation and primary therapy, respectively. Infectious etiologies respond to antimicrobial medications or treatment of the underlying immunosuppressive state. Finally, eosinophilic esophagitis is an important and increasingly recognized cause of dysphagia that is amenable to treatment by elimination of dietary allergens, proton pump inhibition or swallowed, topically acting glucocorticoids in combination with esophageal dilation for persistent strictures.

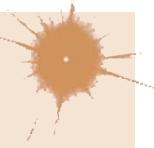
FURTHER READING

- HIRANO I: Esophagus: Anatomy and structural anomalies, in *Yamada Atlas of Gastroenterology*, 6th ed. New York, Wiley-Blackwell Publishing Co., 2016, pp 42–59.
- KAHRILAS PJ et al: The Chicago Classification of esophageal motility disorders, v3.0. *Neurogastroenterol Motil* 27:160, 2015.
- KIM JP, KAHRILAS PJ: How I approach dysphagia. *Curr Gastroenterol Rep* 21:49, 2019.
- PANDOLFINO JP, KAHRILAS PJ: Esophageal neuromuscular function and motility disorders, in *Sleisenger and Fordtran's Gastrointestinal and Liver Disease*, 10th ed, Feldman M, Friedman LS, Brandt LJ (eds). Philadelphia, Elsevier, 2016, pp 701–732.
- SHAKER R et al (eds): *Principles of Deglutition: A Multidisciplinary Text for Swallowing and Its Disorders*. New York, Springer, 2013.

45

Nausea, Vomiting, and Indigestion

William L. Hasler



Nausea is the feeling of a need to vomit. **Vomiting** (emesis) is the oral expulsion of gastrointestinal contents resulting from gut and thoraco-abdominal wall contractions. Vomiting is contrasted with **regurgitation**, the effortless passage of gastric contents into the mouth. **Rumination** is the repeated regurgitation of food residue, which may be rechewed and reswallowed. In contrast to emesis, these phenomena exhibit volitional control. **Indigestion** encompasses a range of complaints including nausea, vomiting, heartburn, regurgitation, and dyspepsia (symptoms thought to originate in the gastroduodenal region). Some individuals with dyspepsia experience postprandial fullness, early satiety (inability to complete a meal due to premature fullness), bloating, eructation (belching), and anorexia. Others report predominantly epigastric burning or pain. Nausea, vomiting, and dyspepsia have been correlated with a condition now called avoidant/restrictive food intake disorder.

NAUSEA AND VOMITING

MECHANISMS

Vomiting is coordinated by the brainstem and is effected by responses in the gut, pharynx, and somatic musculature. Mechanisms underlying nausea are poorly understood but likely involve the cerebral cortex, as nausea requires cognitive and emotional input and is associated with autonomic responses including diaphoresis, pallor, and altered heart rate. Functional brain imaging studies support this idea showing activation of cerebral regions including the insula, anterior cingulate cortex, and amygdala during nausea.

Coordination of Emesis Brainstem nuclei—including the nucleus tractus solitarius; dorsal vagal and phrenic nuclei; medullary nuclei regulating respiration; and nuclei that control pharyngeal, facial, and tongue movements—coordinate initiation of emesis involving neurokinin NK₁, serotonin 5-HT₃, endocannabinoid, and vasopressin pathways.

Somatic and visceral muscles respond stereotypically during emesis. Inspiratory thoracic and abdominal wall muscles contract, increasing intrathoracic and intraabdominal pressures to evacuate the stomach. Under normal conditions, distally migrating gut contractions are coordinated by an electrical phenomenon, the slow wave, which cycles at 3 cycles/min in the stomach and 11 cycles/min in the duodenum. During emesis, slow waves are abolished and replaced by orally propagating spikes that evoke retrograde contractions to facilitate expulsion of gut contents.

Activators of Emesis Emetic stimuli act at several sites. Emesis evoked by unpleasant thoughts or smells originates in the brain. Motion sickness and inner ear disorders act on labyrinthine pathways. Gastric irritants and cytotoxic agents like cisplatin stimulate gastroduodenal vagal afferent nerves. Nongastric afferents are activated by bowel obstruction and mesenteric ischemia. The area postrema, in the medulla, responds to bloodborne stimuli (emetogenic drugs, bacterial toxins, uremia, hypoxia, ketoacidosis) and is termed the *chemoreceptor trigger zone*.

Neurotransmitters mediating vomiting are selective for different sites. Labyrinthine disorders stimulate vestibular muscarinic M₁ and histaminergic H₁ receptors. Vagal afferent stimuli activate 5-HT₃ receptors. The area postrema is served by nerves acting on 5-HT₃, M₁, H₁, and dopamine D₂ subtypes. NK₁ receptors in the central nervous system (CNS) mediate both nausea and vomiting. Cannabinoid CB₁ pathways may participate in the cerebral cortex and brainstem. Therapies for vomiting act on these receptor-mediated pathways.

TABLE 45-1 Causes of Nausea and Vomiting

INTRAPERITONEAL	EXTRAPERITONEAL	MEDICATIONS/ METABOLIC DISORDERS
Obstructing disorders	Cardiopulmonary disease	Drugs
Pyloric obstruction	Cardiomyopathy	Cancer chemotherapy
Small-bowel obstruction	Myocardial infarction	Analgesics
Colonic obstruction	Labyrinthine disease	Opioids
Superior mesenteric artery syndrome	Motion sickness	Antibiotics
Enteric infections	Labyrinthitis	Cardiac antiarrhythmics
Viral	Malignancy	Digoxin
Bacterial	Intracerebral disorders	Oral hypoglycemics
Inflammatory diseases	Malignancy	Oral contraceptives
Cholecystitis	Hemorrhage	Antidepressants
Pancreatitis	Abscess	Restless legs/ Parkinson's therapies
Appendicitis	Hydrocephalus	Smoking cessation agents
Hepatitis	Psychiatric illness	Endocrine/metabolic disease
Altered sensorimotor function	Anorexia and bulimia nervosa	Pregnancy
Gastroparesis	Depression	Uremia
Intestinal pseudoobstruction	Postoperative vomiting	Ketoacidosis
Gastroesophageal reflux		Thyroid and parathyroid disease
Chronic nausea vomiting syndrome		Adrenal insufficiency
Cyclic vomiting syndrome		Toxins
Cannabinoid hyperemesis syndrome		Liver failure
Rumination syndrome		Ethanol
Mesenteric insufficiency		
Celiac artery stenosis		
Median arcuate ligament syndrome		
Biliary colic		
Abdominal irradiation		

■ DIFFERENTIAL DIAGNOSIS

Nausea and vomiting are caused by conditions within and outside the gut, drugs, and circulating toxins (Table 45-1). Unexplained chronic nausea and vomiting is reported by 2–3% of the population.

Intraperitoneal Disorders Obstruction and inflammation of hollow and solid viscera may elicit vomiting. Ulcers and malignancy cause gastric obstruction, while adhesions, benign or malignant tumors, volvulus, intussusception, or inflammatory diseases like Crohn's disease cause small intestinal and colonic obstruction. The superior mesenteric artery syndrome, occurring after weight loss or prolonged bed rest, results when the duodenum is compressed by the overlying superior mesenteric artery. Median arcuate ligament syndrome, with compression of the celiac artery, is a rare cause of vomiting. Abdominal irradiation impairs intestinal motility and induces strictures. Biliary colic causes nausea by acting on afferent nerves. Vomiting with pancreatitis, cholecystitis, and appendicitis results from visceral irritation and induction of ileus. Enteric infectious causes of vomiting include viruses (norovirus, rotavirus), bacteria (*Staphylococcus aureus*, *Bacillus cereus*), and opportunistic organisms like cytomegalovirus or herpes simplex in immunocompromised individuals.

Gut sensorimotor dysfunction often causes nausea and vomiting. *Gastroparesis* presents with these symptoms with evidence of delayed gastric emptying and occurs after vagotomy or with pancreatic carcinoma, mesenteric vascular insufficiency, or organic diseases like diabetes, scleroderma, and amyloidosis. Idiopathic gastroparesis is the most prevalent etiology; it occurs in the absence of systemic illness and follows a viral illness in ~15–20% of cases. Rapid gastric emptying

is associated with nausea and vomiting in some conditions. *Intestinal pseudoobstruction* is characterized by disrupted intestinal motility with retention of food residue and secretions; bacterial overgrowth; nutrient malabsorption; and symptoms of nausea, vomiting, bloating, pain, and altered defecation. Intestinal pseudoobstruction may be idiopathic, inherited, result from systemic disease like scleroderma or an infiltrative process like amyloidosis, or occur as a paraneoplastic consequence of malignancy (e.g., small-cell lung carcinoma). Patients with gastroesophageal reflux, irritable bowel syndrome (IBS), or chronic constipation often report nausea and vomiting.

Other functional gastroduodenal disorders without organic abnormalities have been characterized. *Chronic nausea vomiting syndrome* is defined as bothersome nausea at least 1 day and/or one or more vomiting episodes weekly in the absence of an eating disorder or psychiatric disease. *Cyclic vomiting syndrome (CVS)* causes 3–14% of cases of unexplained nausea and vomiting and presents with discrete episodes of relentless vomiting and is associated with migraines. Some adult cases have been associated with rapid gastric emptying. A related condition, *cannabinoid hyperemesis syndrome (CHS)*, presents with cyclical vomiting in individuals (mostly men) with long-standing use of large quantities of cannabis and resolves with its discontinuation. *Rumination syndrome* is often misdiagnosed as refractory vomiting.

Extraperitoneal Disorders Myocardial infarction and congestive heart failure may cause nausea and vomiting. Postoperative emesis occurs after 25% of surgeries, especially abdominal and orthopedic surgery. Increased intracranial pressure from tumors, bleeding, abscess, or blockage of cerebrospinal fluid outflow produces vomiting with or without nausea. Patients with anorexia nervosa, bulimia nervosa, anxiety, and depression often report significant nausea associated with delayed gastric emptying.

Medications and Metabolic Disorders Drugs evoke vomiting by action on the stomach (analgesics, erythromycin) or area postrema (opioids, anti-parkinsonian drugs). Other emetogenic agents include antibiotics, cardiac antiarrhythmics, antihypertensives, oral hypoglycemics, antidepressants (selective serotonin and serotonin norepinephrine reuptake inhibitors), smoking cessation drugs (varenicline, nicotine), and contraceptives. Cancer chemotherapy causes acute (within hours of administration), delayed (after 1 or more days), or anticipatory vomiting. Acute emesis from highly emetogenic agents (e.g., cisplatin) is mediated by 5-HT₃ pathways. Delayed emesis is more dependent on NK₁ mechanisms. Anticipatory nausea may respond to anxiolytic therapy rather than antiemetics.

Metabolic disorders elicit nausea and vomiting. Nausea affects 70% of women in the first trimester of pregnancy. Hyperemesis gravidarum is a severe form of nausea of pregnancy that produces dehydration and electrolyte disturbances and has been proposed to result from excessive amounts of a blood protein—growth differentiation factor 15. Uremia, ketoacidosis, adrenal insufficiency, and parathyroid and thyroid disease are other metabolic etiologies.

Circulating toxins evoke emesis via effects on the area postrema. Endogenous toxins are generated in fulminant liver failure, whereas exogenous enterotoxins may be produced by enteric bacterial infection. Ethanol intoxication is a common toxic etiology of nausea and vomiting.

APPROACH TO THE PATIENT

Nausea and Vomiting

HISTORY AND PHYSICAL EXAMINATION

The history helps define the etiology of nausea and vomiting. Drugs, toxins, and infections often cause acute symptoms, whereas established illnesses evoke chronic complaints. Gastroparesis and pyloric obstruction elicit vomiting within an hour of eating. Emesis from intestinal blockage occurs later. Vomiting occurring minutes after meal consumption prompts consideration of rumination syndrome. With severe gastric emptying delays, vomitus may contain food residue ingested days before. Hematemesis raises suspicion

of ulcer, malignancy, or Mallory-Weiss tear. Feculent emesis is noted with distal intestinal or colonic obstruction. Bilious vomiting excludes gastric obstruction, whereas emesis of undigested food is consistent with a Zenker's diverticulum or achalasia. Vomiting can relieve abdominal pain from a bowel obstruction but has no effect in pancreatitis or cholecystitis. Weight loss raises concern about malignancy. Taking prolonged hot baths or showers is associated with CHS and CVS. Intracranial sources are considered if there are headaches or visual changes. Vertigo or tinnitus indicates labyrinthine disease.

The physical examination complements the history. Orthostatic hypotension and reduced skin turgor indicate intravascular fluid loss. Pulmonary abnormalities raise concern for aspiration of vomitus. Bowel sounds are absent with ileus. High-pitched rushes suggest bowel obstruction, whereas a succussion splash is found with gastroparesis or pyloric obstruction. Involuntary guarding raises suspicion of inflammation. Fecal blood suggests ulcer, ischemia, or tumor. Neurologic disease presents with papilledema, visual loss, or focal neural abnormalities. Neoplasm is suggested by palpable masses or adenopathy.

DIAGNOSTIC TESTING

For intractable symptoms or an elusive diagnosis, screening testing can direct care. Electrolyte replacement is indicated for hypokalemia or metabolic alkalosis. Iron-deficiency anemia mandates exclusion of mucosal causes. Abnormal pancreatic or liver biochemistries are found with pancreaticobiliary disease. Endocrinologic, rheumatologic, or paraneoplastic etiologies are suggested by hormone or serologic abnormalities. Supine and upright abdominal radiographs may show intestinal air-fluid levels and reduced colonic air with small-bowel obstruction. Ileus is characterized by diffusely dilated air-filled bowel loops.

Anatomic studies are indicated if initial testing is nondiagnostic. Upper endoscopy detects ulcers, malignancy, and retained food in gastroparesis. Small-bowel barium radiography or computed tomography (CT) diagnoses partial bowel obstruction. Colonoscopy or contrast enema radiography detects colonic obstruction. Ultrasound or CT defines intraperitoneal inflammation; CT and magnetic resonance imaging (MRI) enterography define inflammation in Crohn's disease. Brain CT or MRI delineates intracranial disease. Mesenteric angiography, CT, or MRI is useful for suspected ischemia.

Gastrointestinal motility testing can detect an underlying motor disorder. Gastroparesis commonly is diagnosed by gastric scintigraphy, which measures emptying of a radiolabeled meal. A nonradioactive ¹³C-labeled gastric emptying breath test is an alternative to scintigraphy. Intestinal pseudoobstruction is suggested by luminal dilation on imaging or abnormal transit on contrast radiography or intestinal scintigraphy. Wireless motility capsules diagnose gastroparesis or small-bowel dysmotility by detecting local or generalized transit delays in the stomach or small bowel from characteristic pH changes between regions. Small-intestinal manometry confirms a diagnosis of pseudoobstruction and discriminates between neuropathic or myopathic disease based on contractile patterns. Manometry can obviate the need for surgical intestinal biopsy to detect smooth muscle or neuronal degeneration. Combined ambulatory esophageal pH/impedance testing and high-resolution manometry facilitates diagnosis of rumination syndrome. Impedance planimetry detects reduced pyloric distensibility in some cases of gastroparesis.

TREATMENT

Nausea and Vomiting

GENERAL PRINCIPLES

Therapy of vomiting is tailored to correct remediable abnormalities if possible. Patients with severe dehydration should be hospitalized if oral fluid replenishment is unsustainable. Once oral intake is

tolerated, low-fat liquid nutrients are restarted because lipids delay gastric emptying. Low-residue, small-particle diets have shown efficacy in gastroparesis. Glycemic control should be optimized to reduce diabetic gastroparesis symptoms.

ANTIEMETIC MEDICATIONS

Most antiemetic agents act on CNS sites (Table 45-2). Antihistamines like dimenhydrinate and meclizine and anticholinergics like scopolamine act on vestibular pathways to treat motion sickness and labyrinthine disorders. D₂ antagonists treat emesis evoked by area postrema stimuli including medications, toxins, and metabolic disturbances. Dopamine antagonists cross the blood-brain barrier and cause anxiety, movement disorders, and hyperprolactinemic effects (galactorrhea, sexual dysfunction).

Other classes exhibit antiemetic properties. 5-HT₃ antagonists like ondansetron and granisetron prevent postoperative vomiting, radiation therapy-induced symptoms, and cancer chemotherapy-induced emesis, but also are used for other conditions. NK₁ antagonists like aprepitant are approved for chemotherapy-induced vomiting. Aprepitant reduces gastroparesis symptoms. Tricyclic antidepressants reduce symptoms in some patients with functional causes of vomiting, but did not show benefits in a controlled trial in gastroparesis. Other antidepressants such as mirtazapine and olanzapine and the pain-modulating agent gabapentin also exhibit antiemetic effects in some clinical settings.

GASTROINTESTINAL MOTOR STIMULANTS

Drugs that stimulate gastric emptying are used for gastroparesis (Table 45-2). Metoclopramide, a combined 5-HT₄ agonist and D₂ antagonist, is effective in gastroparesis, but antidopaminergic side effects, including dystonias and mood disturbances, limit use in ~25% of cases. Erythromycin increases gastroduodenal motility by action on receptors for motilin, an endogenous transmitter that regulates fasting motility. Intravenous erythromycin is useful for inpatients with refractory gastroparesis. Benefits of long-term oral erythromycin are limited by development of tolerance. Domperidone, a D₂ antagonist not available in the United States, exhibits prokinetic and antiemetic effects but does not cross into most brain regions. The drug rarely causes dystonic reactions but can induce hyperprolactinemic side effects via penetration of pituitary regions served by a porous blood-brain barrier. Prucalopride, a 5-HT₄ agonist, has shown efficacy in accelerating gastric emptying and improving symptoms in idiopathic gastroparesis.

Refractory motility disorders pose challenges. Intestinal pseudodysfunction may respond to the somatostatin analogue octreotide, which induces propagative small-intestinal motor complexes. Acetylcholinesterase inhibitors like pyridostigmine benefit some patients with small-bowel dysmotility. Pyloric botulinum toxin injections reduced gastroparesis symptoms in uncontrolled studies, but small controlled trials observed benefits no greater than sham treatments. Surgical pyloroplasty and gastric peroral endoscopic myotomy (G-POEM) of the pylorus improved symptoms in case series. Enteral feedings through a jejunostomy reduce hospitalizations and improve overall health in some patients with refractory gastroparesis. Subtotal gastric resection may improve some cases of postvagotomy gastroparesis, but its utility for other gastroparesis etiologies is unproven. Implanted gastric electrical stimulators may reduce symptoms, enhance nutrition, improve quality of life, and decrease health care expenditures in medication-refractory gastroparesis; a controlled trial has confirmed modest improvement in vomiting.

SAFETY CONSIDERATIONS

Safety concerns have been raised about selected antiemetics. Metoclopramide can cause irreversible movement disorders like tardive dyskinesia, particularly in older patients. This complication should be explained and documented in the medical record. Domperidone, erythromycin, tricyclic antidepressants, and 5-HT₃ antagonists increase risk of cardiac arrhythmias and sudden cardiac death in

TABLE 45-2 Treatment of Nausea and Vomiting

TREATMENT	MECHANISM	EXAMPLES	CLINICAL INDICATIONS
Antiemetic agents	Antihistaminergic	Dimenhydrinate, meclizine	Motion sickness, inner ear disease
	Anticholinergic	Scopolamine	Motion sickness, inner ear disease
	Antidopaminergic	Prochlorperazine, thiethylperazine, haloperidol	Medication-, toxin-, or metabolic-induced emesis, chemotherapy-induced nausea and vomiting, ?cannabinoid hyperemesis syndrome
	5-HT ₃ antagonist	Ondansetron, granisetron	Chemotherapy- and radiation-induced emesis, postoperative emesis, opioid-induced nausea and vomiting
	Cannabinoids	Tetrahydrocannabinol	Chemotherapy-induced emesis
	Tricyclic antidepressant	Amitriptyline, nortriptyline	Functional vomiting, chronic idiopathic nausea, cyclic vomiting syndrome, ?gastroparesis
	Other antidepressant	Mirtazapine, olanzapine	Functional dyspepsia, ?gastroparesis
	Neuropathic modulator	Gabapentin	Chemotherapy-induced nausea and vomiting
	Neurokinin (NK1) receptor antagonists	Aprepitant, fosaprepitant, netupitant, rolapitant	Chemotherapy-induced emesis
Prokinetic agents	5-HT ₄ agonist and antidopaminergic	Metoclopramide	Gastroparesis
	Motilin agonist	Erythromycin	Gastroparesis, ?intestinal pseudoobstruction
	Peripheral antidopaminergic	Domperidone	Gastroparesis
	Pure 5-HT ₄ agonist	Prucalopride	?idiopathic gastroparesis
	Somatostatin analogue	Octreotide	Intestinal pseudoobstruction
	Acetylcholinesterase inhibitor	Pyridostigmine	?Small-intestinal dysmotility/pseudoobstruction
Special settings	Benzodiazepines	Lorazepam	Anticipatory nausea and vomiting with chemotherapy, cyclic vomiting syndrome
	5-HT _{1A} agonist	Buspirone	Functional dyspepsia
	Glucocorticoids	Methylprednisolone, dexamethasone	Chemotherapy-induced emesis
	Anticonvulsants	Topiramate, zonisamide, levetiracetam	Cyclic vomiting syndrome
	Antimigraine agents	Sumatriptan	Cyclic vomiting syndrome
	Topical analgesic	Capsaicin cream	?Cannabinoid hyperemesis syndrome
	Atypical antipsychotic agent	Olanzapine	Chemotherapy-induced and breakthrough emesis

Note: ?, indication is uncertain.

those with QTc interval prolongation on electrocardiography (ECG). Surveillance ECG testing is advocated for some of these agents.

OTHER CLINICAL SETTINGS

Some cancer chemotherapies are intensely emetogenic (Chap. 73). Combining a 5-HT₃ antagonist, an NK₁ antagonist, and a glucocorticoid can control both acute and delayed vomiting after highly emetogenic chemotherapy. Benzodiazepines like lorazepam reduce anticipatory nausea and vomiting. Other therapies with benefit in chemotherapy-induced emesis include cannabinoids, olanzapine, gabapentin, and alternative therapies like ginger. Most antiemetic regimens produce greater reductions in chemotherapy-induced vomiting than nausea.

Clinicians should exercise caution in managing nausea of pregnancy. Studies of the teratogenic effects of antiemetic agents provide conflicting results. Antihistamines like meclizine and doxylamine, antidopaminergics like prochlorperazine, and antiserotonergics like ondansetron demonstrate limited efficacy. Some obstetricians recommend alternative therapies including pyridoxine, acupressure, or ginger.

Managing CVS and CHS is challenging. Prophylaxis with tricyclic antidepressants or anticonvulsants (topiramate, zonisamide, levetiracetam) reduces the severity and frequency of CVS attacks in uncontrolled reports. Combining intravenous 5-HT₃ antagonists with the sedating effects of a benzodiazepine like lorazepam are mainstays for aborting acute flares. Small studies report benefits with aprepitant and injectable or intranasal forms of the 5-HT₁ agonist sumatriptan to manage acute CVS episodes. These treatments are reportedly less effective for CHS, but haloperidol and topical capsaicin cream may reduce acute CHS attacks.

INDIGESTION

MECHANISMS

Several mechanisms may contribute to indigestion, including acid reflux, altered gut motility or sensation, inflammation, and microbial processes.

Gastroesophageal Reflux Gastroesophageal reflux results from many defects. Reduced lower esophageal sphincter (LES) tone causes reflux in scleroderma and pregnancy and may be a factor in some patients without systemic illness. Other cases exhibit frequent transient LES relaxations (TLESRs). Reductions in esophageal body motility or saliva production prolong esophageal fluid clearance. Increased intragastric pressure promotes gastroesophageal reflux with obesity. Many reflux patients have hiatal hernias, and large hernias can increase symptomatic reflux.

Gastric Motor Dysfunction Disturbed gastric motility may contribute to gastroesophageal reflux in up to one-third of cases. Delayed gastric emptying is found in ~30% of functional dyspeptics, while rapid gastric emptying affects 5%. Impaired gastric fundus relaxation after eating (i.e., accommodation) may underlie selected dyspeptic symptoms like bloating, nausea, and early satiety in ~40% of patients and may predispose to TLESRs and acid reflux.

Visceral Afferent Hypersensitivity Disturbed gastric sensation is another pathogenic factor in functional dyspepsia. Approximately 35% of dyspeptic patients note discomfort with fundic distention to lower pressures than in healthy controls. Other individuals with dyspepsia exhibit hypersensitivity to chemical stimulation of the stomach with capsaicin or with duodenal acid or lipid perfusion. Some cases of functional heartburn without increased acid or nonacid reflux exhibit heightened perception of normal esophageal acidity.

Immune Activation Increases in duodenal epithelial permeability in functional dyspepsia may relate to increases in eosinophils and mast cells adjacent to submucosal neurons. Increased activation of these cells is proposed to contribute to gastric emptying delays and altered sensory function in functional dyspepsia and may selectively elicit early satiety and epigastric pain. Proliferations in duodenal bacteria were shown to correlate with meal-induced symptoms in functional dyspepsia, suggesting a role for microbiome alterations. Intestinal bile salt release also is proposed to worsen dyspeptic symptoms after

eating. Both dysbiosis and bile may contribute to mucosal permeability defects.

Other Factors *Helicobacter pylori* has a proven etiologic role in peptic ulcer disease but is a minor factor in the genesis of functional dyspepsia. Anxiety and depression may play contributing roles in some functional dyspepsia cases. Functional MRI studies show increased activation of several brain regions, emphasizing CNS contributions. Up to 20% of functional dyspepsia patients report symptom onset after a viral illness, suggesting an infectious trigger. Analgesics cause dyspepsia, whereas nitrates, calcium channel blockers, theophylline, and progesterone promote gastroesophageal reflux. Ethanol, tobacco, and caffeine induce LES relaxation and reflux. Genetic factors predispose to development of reflux and dyspepsia in some cases.

■ DIFFERENTIAL DIAGNOSIS

Gastroesophageal Reflux Disease Heartburn or regurgitation is reported weekly by 18–28% of the population, highlighting the prevalence of gastroesophageal reflux disease (GERD). Most cases of heartburn result from excess acid reflux, but reflux of weakly acidic or nonacidic fluid can produce similar symptoms. Alkaline reflux esophagitis elicits GERD symptoms in patients who have had surgery for peptic ulcer disease. Ten percent of patients with heartburn exhibit no acidic or nonacidic esophageal reflux and are considered to have functional heartburn.

Functional Dyspepsia Approximately 20% of the populace has dyspepsia at least six times yearly, but only 10–20% present to clinicians. Functional dyspepsia, the cause of symptoms in 70–80% of dyspeptic patients, is defined as bothersome postprandial fullness, early satiety, or epigastric pain or burning with symptom onset ≥6 months before diagnosis in the absence of organic cause. Functional dyspepsia is subdivided into postprandial distress syndrome (61% of cases), characterized by meal-induced fullness and early satiety, and epigastric pain syndrome (18% of cases), with epigastric pain or burning that may or may not be meal related. Twenty-one percent of individuals present with overlapping postprandial distress and epigastric pain syndromes. Functional dyspepsia is associated with other functional gut disorders including irritable bowel syndrome and nongastrointestinal disorders like fibromyalgia, chronic fatigue, and anxiety. Most cases follow a benign course, but some with *H. pylori* infection or on nonsteroidal anti-inflammatory drugs (NSAIDs) develop ulcers.

Ulcer Disease Most GERD patients do not exhibit esophageal injury, but 5% develop esophageal ulcers. Symptoms cannot distinguish nonerosive from erosive or ulcerative esophagitis. A minority of cases of dyspepsia stem from gastric or duodenal ulcers. The most common causes of ulcers are *H. pylori* infection and NSAID use. Other rare causes of gastroduodenal ulcers include Crohn's disease (Chap. 326) and Zollinger-Ellison syndrome (Chap. 324), resulting from gastrin overproduction by an endocrine tumor.

Malignancy Dyspeptic patients may seek care because of fear of cancer, but few cases result from malignancy. Esophageal squamous cell carcinoma occurs most often with long-standing tobacco or ethanol intake. Other risks include prior caustic ingestion, achalasia, and the hereditary disorder tylosis. Esophageal adenocarcinoma usually complicates prolonged acid reflux. Eight to 20% of GERD patients exhibit esophageal intestinal metaplasia, termed *Barrett's metaplasia*, which predisposes to esophageal adenocarcinoma (Chap. 80). Gastric malignancies include adenocarcinoma, which is prevalent in certain Asian societies, and lymphoma.

Other Causes Opportunistic fungal or viral esophageal infections may produce heartburn but more often cause odynophagia. Other causes of esophageal inflammation include eosinophilic esophagitis and pill esophagitis. Biliary colic is a potential cause unexplained upper abdominal pain, but most patients report discrete acute episodes of right upper quadrant or epigastric pain rather than chronic burning or fullness. Twenty percent of gastroparesis patients note a predominance

of pain rather than nausea and vomiting. Intestinal lactase deficiency may cause gas, bloating, and discomfort and occurs more commonly in blacks and Asians. Intolerance of other carbohydrates (e.g., fructose, sorbitol) produces similar symptoms. Small-intestinal bacterial overgrowth may cause dyspepsia, as well as bowel dysfunction, distension, and malabsorption. Celiac disease, nonceliac gluten sensitivity, pancreatic disease (chronic pancreatitis, malignancy), hepatocellular carcinoma, Ménétrier's disease, infiltrative diseases (sarcoidosis, mastocytosis, eosinophilic gastroenteritis), mesenteric ischemia, thyroid and parathyroid disease, and abdominal wall strain cause dyspepsia. Extraperitoneal etiologies of indigestion include congestive heart failure and tuberculosis.

APPROACH TO THE PATIENT

Indigestion

HISTORY AND PHYSICAL EXAMINATION

Managing indigestion requires a thorough interview. GERD classically produces heartburn, a substernal warmth that moves toward the neck. Heartburn often is exacerbated by meals and may awaken the patient. Associated symptoms include regurgitation of acid or nonacidic fluid and water brash, the reflex release of salty saliva into the mouth. Atypical symptoms include pharyngitis, asthma, cough, bronchitis, hoarseness, and chest pain that mimics angina. Some patients with acid reflux on esophageal pH testing note abdominal pain instead of heartburn.

Dyspeptic patients report symptoms referable to the upper abdomen that may be meal-related (postprandial distress syndrome) or independent of food ingestion (epigastric pain syndrome). The history in functional dyspepsia may also report symptoms of GERD, IBS, or idiopathic gastroparesis.

The physical exam with GERD and functional dyspepsia usually is normal. In atypical GERD, pharyngeal erythema and wheezing may be noted. Recurrent regurgitation may cause poor dentition. Dyspeptics may exhibit epigastric tenderness or distention.

Discriminating functional from organic causes of indigestion mandates excluding certain historic and exam features. Odynophagia suggests esophageal infection. Dysphagia is concerning for a benign or malignant esophageal blockage. Other alarm features include unexplained weight loss, recurrent vomiting, dysphagia, occult or gross bleeding, nocturnal symptoms, jaundice, palpable mass or adenopathy, and a family history of gastrointestinal neoplasm. Patients with an abdominal wall source of upper abdominal pain may exhibit a positive Carnett's sign of increased tenderness with tensing of abdominal muscles upon lifting the head from the exam table.

DIAGNOSTIC TESTING

Because indigestion is prevalent and most cases result from GERD or functional dyspepsia, it is generally recommended to perform no more than limited and directed diagnostic testing in most individuals.

After excluding alarm factors (Table 45-3), patients with typical GERD do not need further evaluation and are treated empirically. Upper endoscopy is indicated only in cases with atypical symptoms or these alarm factors. For heartburn >5 years in duration,

TABLE 45-3 Alarm Symptoms in Gastroesophageal Reflux Disease

Odynophagia or dysphagia
Unexplained weight loss
Recurrent vomiting
Occult or gross gastrointestinal bleeding
Jaundice
Palpable mass or adenopathy
Family history of gastroesophageal malignancy

especially in patients >50 years old, endoscopy is advocated to screen for Barrett's metaplasia. Endoscopy is not needed in low-risk patients who respond to acid suppressants. Ambulatory esophageal pH testing using a catheter method or a wireless capsule endoscopically attached to the esophageal wall is considered for drug-refractory symptoms and atypical symptoms like unexplained chest pain. High-resolution esophageal manometry is ordered when surgical treatment of GERD is considered. A low LES pressure predicts failure of drug therapy and provides a rationale to proceed to surgery. Poor esophageal body peristalsis raises concern about postoperative dysphagia and directs the choice of surgical technique. Nonacidic reflux may be detected by combined esophageal impedance-pH testing in medication-unresponsive patients.

Upper endoscopy is recommended as the initial test in patients with unexplained dyspepsia who are >60 years old to exclude malignancy—a finding in only 0.3% of endoscopies performed for uninvestigated dyspepsia. Management of patients <60 years old depends on the local *H. pylori* prevalence. In regions with low prevalence (<10%), a 4-week trial of an acid-suppressing medication such as a proton pump inhibitor (PPI) is recommended. If empiric acid suppression fails, a “test and treat” approach for *H. pylori* status is initiated with urea breath testing or stool antigen measurement. Those who are *H. pylori* positive are given therapy to eradicate infection. For patients in areas with high *H. pylori* prevalence (>10%), an initial “test and treat” approach is advocated, and empiric PPI therapy is reserved for those who are negative for infection or who fail to respond to *H. pylori* treatment. Patients who are treated for *H. pylori* should undergo confirmation of eradication with repeat urea breath testing or fecal antigen testing 4–6 weeks after completing therapy. Those under age 60 only warrant upper endoscopy if their symptoms fail to respond to these therapies. Some advocate initial endoscopy for patients <60 years old who report alarm symptoms, but some guidelines have not endorsed this practice unless symptoms persist despite treatment.

Further testing is indicated in some settings. For suspected bleeding, a blood count can exclude anemia. Thyroid chemistries or calcium levels screen for metabolic disease. Specific serologies may suggest celiac disease. Pancreatic and liver chemistries are obtained for suspected pancreaticobiliary causes, which are further investigated with ultrasound, CT, or MRI. Gastric emptying testing is considered to exclude gastroparesis for dyspeptic symptoms resembling postprandial distress when therapy fails. Breath testing after carbohydrate ingestion detects lactase deficiency, intolerance to other carbohydrates, or small-intestinal bacterial overgrowth.

TREATMENT

Indigestion

LIFESTYLE, DIET, AND NONMEDICATION RECOMMENDATIONS

Patients with mild indigestion can be reassured that a careful evaluation revealed no serious disease and are offered no other intervention. If possible, drugs that cause gastroesophageal reflux or dyspepsia should be stopped. GERD patients should limit ethanol, caffeine, chocolate, and tobacco use and can ingest a low-fat diet, avoid snacks before bedtime, and elevate the head of the bed. Functional dyspepsia patients can be advised to reduce intake of fat, spicy foods, caffeine, and alcohol. Dietary lactose restriction is appropriate for lactase deficiency, while gluten exclusion is indicated for celiac disease. Low FODMAP (fermentable oligosaccharide, disaccharide, monosaccharide, and polyol) diets are effective for gaseous symptoms in IBS. In a systematic review, FODMAP intake correlated with functional dyspepsia symptoms, suggesting potential utility in this disorder as well.

ACID-SUPPRESSING OR -NEUTRALIZING MEDICATIONS

Drugs that reduce or neutralize gastric acid are often prescribed for GERD. Histamine H₂ antagonists like cimetidine, ranitidine, famotidine, and nizatidine are useful in mild to moderate GERD. For severe symptoms or for many cases of erosive or ulcerative esophagitis, PPIs like omeprazole, lansoprazole, rabeprazole, pantoprazole, esomeprazole, or dexlansoprazole are needed. These drugs inhibit gastric H⁺, K⁺-ATPase and are more potent than H₂ antagonists. Up to one-third of GERD patients do not respond to standard PPI doses; one-third of these patients have nonacidic reflux, whereas 10% have persistent acid-related disease. Heartburn responds better to PPI therapy than regurgitation or atypical GERD symptoms. Some individuals respond to doubling of the PPI dose or adding an H₂ antagonist. Complications of long-term PPI therapy include diarrhea (*Clostridium difficile* infection, microscopic colitis), small-intestinal bacterial overgrowth, nutrient deficiency (vitamin B₁₂, iron, calcium), hypomagnesemia, bone demineralization, interstitial nephritis, and impaired medication absorption (clopidogrel). Many patients started on a PPI can be stepped down to an H₂ antagonist or switched to on-demand use.

Acid suppressants also are effective for both the postprandial distress and epigastric pain subtypes of functional dyspepsia. A meta-analysis of 18 controlled trials calculated a risk ratio of 0.88, with a 95% confidence interval of 0.82–0.94, favoring PPI therapy over placebo in functional dyspepsia. H₂ antagonists also improve symptoms in functional dyspepsia, but a guideline has advocated PPIs over H₂ antagonists as first-line therapies for functional dyspepsia. In addition to acid suppression, PPIs may have the additional action of reducing duodenal eosinophil counts in dyspepsia.

Antacids are useful for short-term control of mild GERD but have less benefit in severe cases unless given at high doses that cause side effects (diarrhea and constipation with magnesium- and aluminum-containing agents, respectively). Alginic acid combined with antacids forms a floating barrier to reflux in patients with upright symptoms. Sucralfate, a salt of aluminum hydroxide and sucrose octasulfate that buffers acid and binds pepsin and bile salts, shows efficacy in GERD similar to H₂ antagonists.

HELICOBACTER PYLORI ERADICATION

H. pylori eradication is indicated for peptic ulcer and mucosa-associated lymphoid tissue gastric lymphoma. The benefits of eradication therapy in functional dyspepsia are limited but are statistically significant. A systematic review of 25 controlled trials calculated a pooled risk ratio of 1.24, with a 95% confidence interval of 1.12–1.37, favoring *H. pylori* eradication over placebo. Most drug combinations (Chaps. 163 and 324) include 7–14 days of a PPI with two or three antibiotics with or without bismuth products. *H. pylori* infection is associated with reduced prevalence of GERD. However, eradication of infection does not worsen GERD symptoms. No consensus recommendations regarding *H. pylori* eradication in GERD patients have been offered.

AGENTS THAT MODIFY GASTROINTESTINAL MOTOR ACTIVITY

The γ-aminobutyric acid B (GABA-B) agonist baclofen reduces esophageal exposure to acid and nonacidic fluids by reducing TLESRs by 40%. This drug can be used in patients with refractory acid or nonacid reflux. Several studies have promoted the efficacy of agents that stimulate gastric emptying in functional dyspepsia with 33% relative risk reductions, but publication bias and small sample sizes raise questions about reported benefits of these agents. Some clinicians suggest that patients with the postprandial distress subtype may respond preferentially to such prokinetic drugs. The newer 5-HT₄ agonist prucalopride was reported to reduce symptoms in patients with idiopathic gastroparesis, but no similar studies have been conducted in functional dyspepsia. The 5-HT_{1A} agonists buspirone and tandospirone may improve

some functional dyspepsia symptoms by enhancing meal-induced gastric accommodation. Acotiamide stimulates gastric emptying and augments accommodation by enhancing acetylcholine release via muscarinic receptor antagonism and acetylcholinesterase inhibition. This agent is approved for functional dyspepsia in Japan and India.

ANTIDEPRESSANTS

Some patients with refractory functional heartburn may respond to antidepressants in the tricyclic and selective serotonin reuptake inhibitor (SSRI) classes, although studies are limited. Their mechanism of action may involve blunting of visceral pain processing in the brain. In a controlled trial in functional dyspepsia, the tricyclic drug amitriptyline produced symptom reductions, whereas the SSRI escitalopram had no benefit in a three-way comparison with placebo. In another controlled trial in functional dyspepsia, the antidepressant mirtazapine produced superior symptom reductions versus placebo. However, in a meta-analysis of 13 trials, SSRIs and serotonin-norepinephrine reuptake inhibitors showed no benefits in functional dyspepsia.

OTHER OPTIONS

Antireflux surgery (fundoplication) to enhance the barrier function of the LES may be offered to GERD patients who are young and require lifelong therapy, have typical heartburn, are responsive to PPIs, and show acid reflux on pH monitoring. Surgery also is effective for some cases of nonacidic reflux. Individuals who respond less well to fundoplication include those with atypical symptoms, those who have functional heartburn without reflux on testing, or those who have esophageal body motor disturbances. Dysphagia, gas-bloat syndrome, and gastroparesis are long-term complications of fundoplication; ~60% develop recurrent GERD symptoms over time. Magnetic sphincter augmentation may be appropriate for GERD treatment, while endoscopic radiofrequency therapies can be considered for some patients. Other endoscopic options including transoral incisionless fundoplication, endoscopic stapling, and antireflux mucosectomy are not yet advocated.

Gas and bloating are bothersome in some patients with indigestion and are difficult to treat. Simethicone, activated charcoal, and alpha-galactosidase provide benefits in some cases. One trial suggested possible benefits of the nonabsorbable antibiotic rifaximin in functional dyspepsia, while another reported improvement with the probiotic *Lactobacillus gasseri*. Herbal remedies like STW 5 (Iberogast, a mixture of nine herbal agents) and formulations of caraway oil and menthol are useful in some dyspeptic patients. Psychological treatments (e.g., behavioral therapy, psychotherapy, hypnotherapy) may be offered for refractory functional dyspepsia; a meta-analysis of four trials reported benefits in patients with persistent dyspepsia.

FURTHER READING

- GYAWALI CP et al: ACG Clinical Guidelines: clinical use of esophageal physiologic testing. *Am J Gastroenterol* 115:1412, 2020.
- MARET-OUDA J et al: Gastroesophageal reflux disease: a review. *JAMA* 324:2536, 2020.
- SHARAF RN et al: Management of cyclic vomiting syndrome in adults: evidence review. *Neurogastroenterol Motil* 31(Suppl 2):e13605, 2019.
- VENKATESAN T et al: Role of chronic cannabis use: cyclic vomiting syndrome vs. cannabinoid hyperemesis syndrome. *Neurogastroenterol Motil* 31(Suppl 2):e13606, 2019.
- WAUTERS L et al: Novel concepts in the pathophysiology and treatment of functional dyspepsia. *Gut* 69:591, 2020.

46

Diarrhea and Constipation

Michael Camilleri, Joseph A. Murray



Diarrhea and constipation are exceedingly common and, together, exact an enormous toll in terms of mortality, morbidity, social inconvenience, loss of work productivity, and consumption of medical resources. Worldwide, >1 billion individuals suffer one or more episodes of acute diarrhea each year. Among the 100 million persons affected annually by acute diarrhea in the United States, nearly half must restrict activities, 10% consult physicians, ~250,000 require hospitalization, and ~5000 die (primarily the elderly). Updated 2014–2015 annual disease burden data from the United States show 3.4 million annual clinic or emergency department visits, about 130,000 hospital admissions, and annual economic burden to society (excluding all costs for inflammatory bowel disease) exceeding \$8 billion. Acute infectious diarrhea remains one of the most common causes of mortality in developing countries, particularly among impoverished infants, accounting for 1.8 million deaths per year. Recurrent, acute diarrhea in children in tropical countries results in environmental enteropathy with long-term impacts on physical and intellectual development.

Constipation, by contrast, is rarely associated with mortality and is exceedingly common in developed countries, leading to frequent self-medication and, in a third of those, to medical consultation. Annual disease burden data for 2014–2015 show about 5 million clinic or emergency department visits for constipation or hemorrhoids, 50,000 admissions to hospital, and average cost of \$3500 per patient, about double that of controls in a nested controlled study.

Population statistics on chronic diarrhea and constipation are more uncertain, perhaps due to variable definitions and reporting, but the frequency of these conditions is also high. U.S. population surveys put prevalence rates for chronic diarrhea at 2–7% and for chronic constipation at 12–19%, with women being affected twice as often as men, reaching parity at 70 years of age. Diarrhea and constipation are among the most common patient complaints presenting in primary care and account for nearly 50% of referrals to gastroenterologists.

Although diarrhea and constipation may present as mere nuisance symptoms at one extreme, they can be severe or life threatening at the other. Even mild symptoms may signal a serious underlying gastrointestinal (GI) lesion, such as colorectal cancer, or systemic disorder, such as thyroid disease. Given the heterogeneous causes and potential severity of these common complaints, it is imperative for clinicians to appreciate the pathophysiology, etiologic classification, diagnostic strategies, and principles of management of diarrhea and constipation so that rational and cost-effective care can be delivered.

NORMAL PHYSIOLOGY

While the primary function of the small intestine is the digestion and assimilation of nutrients from food, the small intestine and colon together perform important functions that regulate the secretion and absorption of water and electrolytes, the storage and subsequent transport of intraluminal contents aborally, and the salvage of some nutrients that are not absorbed in the small intestine after bacterial metabolism of carbohydrate allows salvage of short-chain fatty acids. The main motor functions are summarized in **Table 46-1**. Alterations in fluid and electrolyte handling contribute significantly to diarrhea. Alterations in motor and sensory functions of the colon result in highly prevalent syndromes such as irritable bowel syndrome (IBS), chronic diarrhea, and chronic constipation.

NEURAL CONTROL

The small intestine and colon have intrinsic and extrinsic innervation. The *intrinsic innervation*, also called the enteric nervous system, comprises myenteric, submucosal, and mucosal neuronal layers. The function of these layers is modulated by interneurons through the actions

TABLE 46-1 Normal Gastrointestinal Motility: Functions at Different Anatomic Levels**Stomach and Small Bowel**

Synchronized MMC in fasting
Accommodation, trituration, mixing, transit
Stomach ~3 h
Small bowel ~3 h
Ileal reservoir empties boluses

Colon: Irregular Mixing, Fermentation, Absorption, Transit

Ascending, transverse: reservoirs
Descending: conduit
Sigmoid/rectum: volitional reservoir

Abbreviation: MMC, migrating motor complex.

of neurotransmitter amines or peptides, including acetylcholine, vasoactive intestinal peptide (VIP), opioids, norepinephrine, serotonin, adenosine triphosphate (ATP), and nitric oxide (NO). The myenteric plexus regulates smooth-muscle function through intermediary pacemaker-like cells called the interstitial cells of Cajal, and the submucosal plexus affects secretion, absorption, and mucosal blood flow. The enteric nervous system receives input from the extrinsic nerves, but it is capable of independent control of these functions.

The *extrinsic innervations* of the small intestine and colon are part of the autonomic nervous system and also modulate motor and secretory functions. The parasympathetic nerves convey visceral sensory pathways from and excitatory pathways to the small intestine and colon. Parasympathetic fibers via the vagus nerve reach the small intestine and proximal colon along the branches of the superior mesenteric artery. The distal colon is supplied by sacral parasympathetic nerves (S_{2-4}) via the pelvic plexus; these fibers course through the wall of the colon as ascending intracolonic fibers as far as, and in some instances including, the proximal colon. The chief excitatory

neurotransmitters controlling motor function are acetylcholine and the tachykinins, such as substance P. The sympathetic nerve supply modulates motor functions and reaches the small intestine and colon alongside their arterial vessels. Sympathetic input to the gut is generally excitatory to sphincters and inhibitory to nonsphincteric muscle. Visceral afferents convey sensation from the gut to the central nervous system (CNS). Some afferent fibers synapse in the prevertebral ganglia and reflexly modulate intestinal motility, blood flow, and secretion.

■ INTESTINAL FLUID ABSORPTION AND SECRETION

On an average day, 9 L of fluid enter the GI tract, ~1 L of residual fluid reaches the colon, and the stool excretion of fluid constitutes about 0.2 L/d. The colon has a large capacitance and functional reserve and may recover up to four times its usual volume of 0.8 L/d, provided the rate of flow permits reabsorption to occur. Thus, the colon can partially compensate for excess fluid delivery to the colon that may result from intestinal absorptive or secretory disorders.

In the small intestine and colon, sodium absorption is predominantly electrogenic (i.e., it can be measured as an ionic current across the membrane because there is not an equivalent loss of a cation from the cell), and uptake takes place at the apical membrane; it is compensated for by the export functions of the basolateral sodium pump. There are several active transport proteins at the apical membrane, especially in the small intestine, whereby sodium ion entry is coupled to monosaccharides (e.g., glucose through the transporter SGLT1, or fructose through GLUT-5). Glucose then exits the basal membrane through a specific transport protein, GLUT-2, creating a glucose concentration and osmotic gradient between the lumen and the intercellular space, drawing water and electrolytes passively from the lumen. Several channels mediate the secretion of chloride ions in diarrheal diseases or in response to medications administered for the treatment of constipation. The diverse ion channels (chloride channels and cystic fibrosis transmembrane regulator), transporters (SGLT1, GLUT-2), and receptors (e.g., guanylate cyclase C receptor) are summarized in **Figure 46-1**.

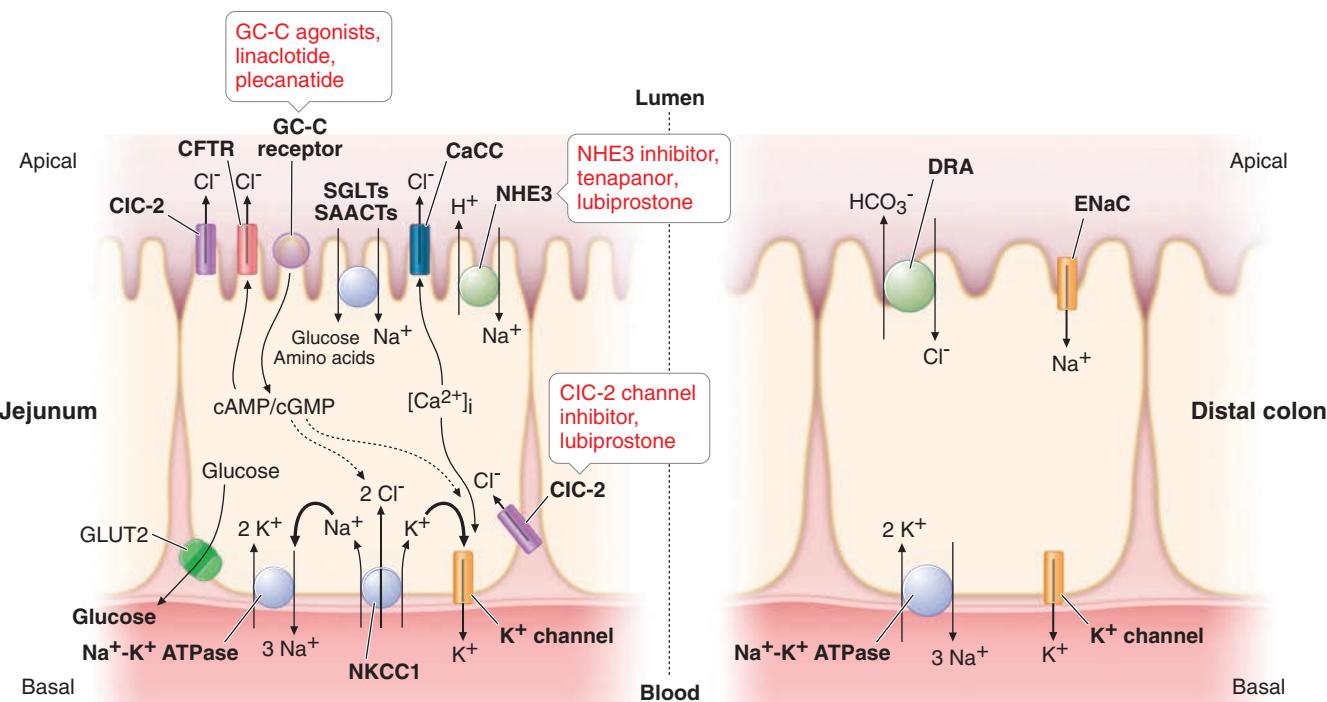


FIGURE 46-1 Important ion transport mechanisms in the jejunum and colon, and the site of action of medications used as secretagogues in the treatment of chronic constipation. CFTR, cystic fibrosis transmembrane regulator; CIC2, type 2 chloride channel, DRA, downregulated in adenoma (also called SLC26A3); ENaC, epithelial sodium channel; GC-C, guanylate cyclase C; Na⁺-K⁺ ATPase, sodium potassium adenosine triphosphatase; NHE3, sodium-hydrogen exchanger; NKCC1, Na-K-Cl cotransporter; SAACCT, sodium amino acid co-transporters; SGLT, sodium glucose transporters.

A variety of neural and nonneuronal mediators regulate colonic fluid and electrolyte balance, including cholinergic, adrenergic, and serotonergic mediators. Angiotensin and aldosterone also influence colonic absorption, reflecting the common embryologic development of the distal colonic epithelium and the renal tubules.

■ SMALL-INTESTINAL MOTILITY

During the fasting period, the motility of the small intestine is characterized by a cyclical event called the migrating motor complex (MMC), which serves to clear nondigestible residue from the small intestine (the intestinal “housekeeper”). This organized, propagated series of contractions lasts, on average, 4 min, occurs every 60–90 min, and usually involves the entire small intestine. After food ingestion, the small intestine produces irregular, mixing contractions of relatively low amplitude, except in the distal ileum where more powerful contractions occur intermittently and empty the ileum by bolus transfers.

■ ILEOCOLONIC STORAGE AND SALVAGE

The distal ileum acts as a reservoir, emptying intermittently by bolus movements. This action allows time for salvage of fluids, electrolytes, and nutrients. Segmentation by haustra compartmentalizes the colon and facilitates mixing, retention of residue, and formation of solid stools. There is increased appreciation of the intimate interaction between the colonic function and the luminal ecology. The resident microorganisms, predominantly anaerobic bacteria, in the colon are necessary for the digestion of unabsorbed carbohydrates that reach the colon even in health, thereby providing a vital source of nutrients to the mucosa. Normal intestinal flora also keeps pathogens at bay by a variety of mechanisms including a crucial role in the development and maintenance of a potent but well-regulated immune response capacity to pathogens and tolerance to normal ingesta. In health, the ascending and transverse regions of colon function as reservoirs (average transit time, 15 h), and the descending colon acts as a conduit (average transit time, 3 h). The colon is efficient at conserving sodium and water, a function that is particularly important in sodium-depleted patients in whom the small intestine alone is unable to maintain sodium balance. Diarrhea or constipation may result from alteration in the reservoir function of the proximal colon or the propulsive function of the left colon. Constipation may also result from disturbances of the rectal or sigmoid reservoir, typically as a result of dysfunction of the pelvic floor, the anal sphincters, the coordination of defecation, or dehydration.

■ COLONIC MOTILITY AND TONE

The small-intestinal MMC only rarely continues into the colon. However, short duration or phasic contractions mix colonic contents, and high-amplitude (>75 mmHg) propagated contractions (HAPCs) are sometimes associated with mass movements through the colon and normally occur approximately five times per day, usually on awakening in the morning and postprandially. Increased frequency of HAPCs may result in diarrhea or urgency. The predominant phasic contractions in the colon are irregular and nonpropagated and serve a “mixing” function.

Colonic tone refers to the background contractility upon which phasic contractile activity (typically contractions lasting <15 s) is superimposed. It is an important cofactor in the colon’s capacitance (volume accommodation) and sensation.

■ COLONIC MOTILITY AFTER MEAL INGESTION

After meal ingestion, colonic phasic and tonic contractility increase for a period of ~2 h. The initial phase (~10 min) is mediated by the vagus

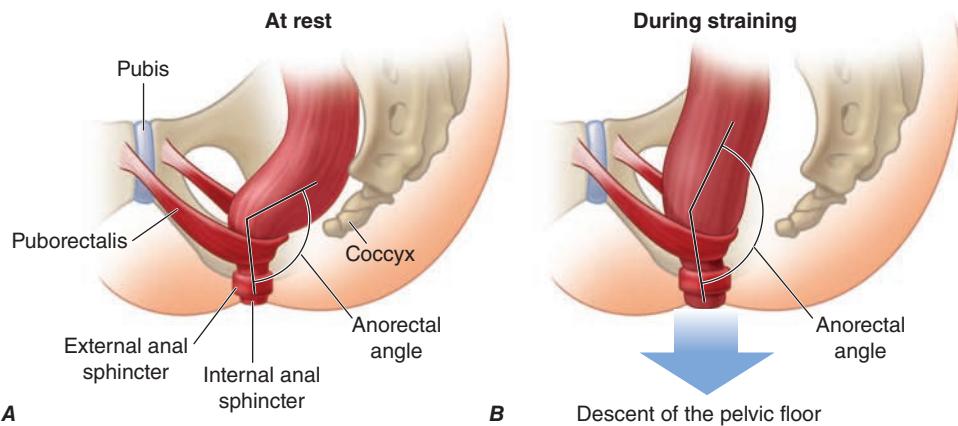


FIGURE 46-2 Sagittal view of the anorectum (**A**) at rest and (**B**) during straining to defecate. Continence is maintained by normal rectal sensation and tonic contraction of the internal anal sphincter and the puborectalis muscle, which wraps around the anorectum, maintaining an anorectal angle between 80° and 110°. During defecation, the pelvic floor muscles (including the puborectalis) relax, allowing the anorectal angle to straighten by at least 15°, and the perineum descends by 1–3.5 cm. The external anal sphincter also relaxes and reduces pressure on the anal canal. (From A Lembo, M Camilleri: Chronic constipation. *N Engl J Med* 349:1360, 2003 Massachusetts Medical Society. Reprinted with permission.)

nerve in response to mechanical distention of the stomach. The subsequent response of the colon requires caloric stimulation (e.g., intake of at least 500 kcal) and is mediated, at least in part, by hormones (e.g., gastrin and serotonin).

■ DEFECATION

Tonic contraction of the puborectalis muscle, which forms a sling around the rectoanal junction, is important to maintain continence; during defecation, sacral parasympathetic nerves relax this muscle, facilitating the straightening of the rectoanal angle (Fig. 46-2). Distention of the rectum results in transient relaxation of the internal anal sphincter via intrinsic and reflex sympathetic innervation. As sigmoid and rectal contractions, as well as straining (Valsalva maneuver), which increases intraabdominal pressure, increase the pressure within the rectum, the rectosigmoid angle opens by >15°. Voluntary relaxation of the external anal sphincter (striated muscle innervated by the pudendal nerve) in response to the sensation produced by distention permits the evacuation of feces. Defecation can also be delayed voluntarily by contraction of the external anal sphincter.

DIARRHEA

■ DEFINITION

Diarrhea is loosely defined as passage of abnormally liquid or unformed stools at an increased frequency. For adults on a typical Western diet, stool weight >200 g/d can generally be considered diarrheal. Diarrhea may be further defined as *acute* if <2 weeks, *persistent* if 2–4 weeks, and *chronic* if >4 weeks in duration.

Two common conditions, usually associated with the passage of stool totaling <200 g/d, must be distinguished from diarrhea, because diagnostic and therapeutic algorithms differ. *Pseudodiarrhea*, or the frequent passage of small volumes of stool, is often associated with rectal urgency, tenesmus, or a feeling of incomplete evacuation and accompanies IBS or proctitis. *Fecal incontinence* is the involuntary discharge of rectal contents and is most often caused by neuromuscular disorders or structural anorectal problems. Diarrhea and urgency, especially if severe, may aggravate or cause incontinence. Pseudodiarrhea and fecal incontinence occur at prevalence rates comparable to or higher than that of chronic diarrhea and should always be considered in patients complaining of “diarrhea.” Overflow diarrhea may occur in nursing home patients due to fecal impaction that is readily detectable by rectal examination. A careful history and physical examination generally allow these conditions to be discriminated from true diarrhea.

■ ACUTE DIARRHEA

More than 90% of cases of acute diarrhea are caused by infectious agents; these cases are often accompanied by vomiting, fever, and

abdominal pain. The remaining 10% or so are caused by medications, toxic ingestions, ischemia, food indiscretions, and other conditions.

Infectious Agents Most infectious diarrheas are acquired by fecal-oral transmission or, more commonly, via ingestion of food or water contaminated with pathogens from human or animal feces. In the immunocompetent person, the resident fecal microflora, containing >500 taxonomically distinct species, are rarely the source of diarrhea and may actually play a role in suppressing the growth of ingested pathogens. Disturbances of flora by antibiotics can lead to diarrhea by reducing the digestive function or by allowing the overgrowth of pathogens, such as *Clostridium difficile* (Chap. 134). Acute infection or injury occurs when the ingested agent overwhelms or bypasses the host's mucosal immune and nonimmune (gastric acid, digestive enzymes, mucus secretion, peristalsis, and suppressive resident flora) defenses. Established clinical associations with specific enteropathogens may offer diagnostic clues. Diarrhea occasionally is an early symptom of infection such as SARS-CoV-2 and *Legionella*.

In the United States, five high-risk groups are recognized:

1. **Travelers.** Nearly 40% of tourists to endemic regions of Latin America, Africa, and Asia develop so-called traveler's diarrhea, most commonly due to enterotoxigenic or enteroaggregative *Escherichia coli* as well as to *Campylobacter*, *Shigella*, *Aeromonas*, norovirus, *Coronavirus*, and *Salmonella*. Visitors to Russia (especially St. Petersburg) may have increased risk of *Giardia*-associated diarrhea; visitors to Nepal may acquire *Cyclospora*. Campers, backpackers, and swimmers in wilderness areas may become infected with *Giardia*. Cruise ships may be affected by outbreaks of gastroenteritis caused by agents such as norovirus.
2. **Consumers of certain foods.** Diarrhea closely following food consumption at a picnic, banquet, or restaurant may suggest infection with *Salmonella*, *Campylobacter*, or *Shigella* from chicken; enterohemorrhagic *E. coli* (O157:H7) from undercooked hamburger; *Bacillus cereus* from fried rice or other reheated food; *Staphylococcus aureus* or *Salmonella* from mayonnaise or creams; *Salmonella* from eggs; *Listeria* from fresh or frozen uncooked foods, mushrooms, or dairy products.

products; and *Vibrio* species, *Salmonella*, or acute hepatitis A from seafood, especially if raw. State departments of public health issue communications regarding domestic and foreign food-related illnesses, often identified by rapid DNA typing (PulseNet), that cause epidemics in the United States (e.g., the *Listeria* epidemic of 2020 from imported enoki mushrooms).

3. **Immunodeficient persons.** Individuals at risk for diarrhea include those with either primary immunodeficiency (e.g., IgA deficiency, common variable hypogammaglobulinemia, chronic granulomatous disease) or the much more common secondary immunodeficiency states (e.g., AIDS, senescence, pharmacologic suppression). Common enteric pathogens often cause a more severe and protracted diarrheal illness, and, particularly in persons with AIDS, opportunistic infections, such as by *Mycobacterium* species, certain viruses (cytomegalovirus, adenovirus, and herpes simplex), and protozoa (*Cryptosporidium*, *Isospora belli*, Microsporidia, and *Blastocystis hominis*) may also play a role (Chap. 202). In patients with AIDS, agents transmitted venereally per rectum or by extension from vaginal infection (e.g., *Neisseria gonorrhoeae*, *Treponema pallidum*, *Chlamydia*) may contribute to proctocolitis. Symptoms suggesting anorectal disease, particularly pain, may result from constipation occurring coincidentally in a person with immunodeficiency. Persons with hemochromatosis are especially prone to invasive, even fatal, enteric infections with *Vibrio* species and *Yersinia* infections and should avoid raw fish and exposing open wounds to seawater.
4. **Daycare attendees and their family members.** Infections with *Shigella*, *Giardia*, *Cryptosporidium*, rotavirus, and other agents are very common and should be considered.
5. **Institutionalized persons.** Infectious diarrhea is one of the most frequent categories of nosocomial infections in many hospitals and long-term care facilities; the causes are a variety of microorganisms but most commonly *C. difficile*. *C. difficile* can affect those with no history of antibiotic use and is often community acquired.

The pathophysiology underlying acute diarrhea by infectious agents produces specific clinical features that may also be helpful in diagnosis (Table 46-2). Profuse, watery diarrhea secondary to small-bowel hypersecretion occurs with ingestion of preformed bacterial toxins,

TABLE 46-2 Association Between Pathobiology of Causative Agents and Clinical Features in Acute Infectious Diarrhea

PATHOBIOLOGY/AGENTS	INCUBATION PERIOD	VOMITING	ABDOMINAL PAIN	FEVER	DIARRHEA
Toxin producers					
Preformed toxin					
<i>Bacillus cereus</i> , <i>Staphylococcus aureus</i> , <i>Clostridium perfringens</i>	1–8 h 8–24 h	3–4+	1–2+	0–1+	3–4+, watery
Enterotoxin					
<i>Vibrio cholerae</i> , enterotoxigenic <i>Escherichia coli</i> , <i>Klebsiella pneumoniae</i> , <i>Aeromonas</i> species	8–72 h	2–4+	1–2+	0–1+	3–4+, watery
Enteroadherent					
Enteropathogenic and enteroadherent <i>E. coli</i> , <i>Giardia</i> organisms, cryptosporidiosis, helminths	1–8 d	0–1+	1–3+	0–2+	1–2+, watery, mushy
Cytotoxin producers					
<i>Clostridium difficile</i>	1–3 d	0–1+	3–4+	1–2+	1–3+, usually watery, occasionally bloody
Hemorrhagic <i>E. coli</i>	12–72 h	0–1+	3–4+	1–2+	1–3+, initially watery, quickly bloody
Invasive organisms					
Minimal inflammation					
Rotavirus and norovirus	1–3 d	1–3+	2–3+	3–4+	1–3+, watery
Variable inflammation					
<i>Salmonella</i> , <i>Campylobacter</i> , and <i>Aeromonas</i> species, <i>Vibrio parahaemolyticus</i> , <i>Yersinia</i>	12 h–11 d	0–3+	2–4+	3–4+	1–4+, watery or bloody
Severe inflammation					
<i>Shigella</i> species, enteroinvasive <i>E. coli</i> , <i>Entamoeba histolytica</i>	12 h–8 d	0–1+	3–4+	3–4+	1–2+, bloody

Source: Adapted from DW Powell, in T Yamada (ed): *Textbook of Gastroenterology and Hepatology*, 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2003.

enterotoxin-producing bacteria, and enteroadherent pathogens. Diarrhea associated with marked vomiting and minimal or no fever may occur abruptly within a few hours after ingestion of the former two types; vomiting is usually less, abdominal cramping or bloating is greater, and fever is higher with the latter. Cytotoxin-producing and invasive microorganisms all cause high fever and abdominal pain. Invasive bacteria and *Entamoeba histolytica* often cause bloody diarrhea (referred to as *dysentery*). *Yersinia* invades the terminal ileal and proximal colon mucosa and may cause especially severe abdominal pain with tenderness mimicking acute appendicitis.

Finally, infectious diarrhea may be associated with systemic manifestations. Reactive arthritis (formerly known as Reiter's syndrome), arthritis, urethritis, and conjunctivitis may accompany or follow infections by *Salmonella*, *Campylobacter*, *Shigella*, and *Yersinia*. Yersiniosis may also lead to an autoimmune-type thyroiditis, pericarditis, and glomerulonephritis. Both enterohemorrhagic *E. coli* (O157:H7) and *Shigella* can lead to the *hemolytic-uremic syndrome* with an attendant high mortality rate. The syndrome of postinfectious IBS has now been recognized as a complication of infectious diarrhea. Similarly, acute gastroenteritis may precede the diagnosis of celiac disease or Crohn's disease. Acute diarrhea can also be a major symptom of several systemic infections including *viral hepatitis*, *listeriosis*, *legionellosis*, and *toxic shock syndrome*.

Other Causes Side effects from medications are probably the most common noninfectious causes of acute diarrhea, and etiology may be suggested by a temporal association between use and symptom onset. Although innumerable medications may produce diarrhea, some of the more frequently incriminated include antibiotics, cardiac antidysrhythmics, antihypertensives, nonsteroidal anti-inflammatory drugs (NSAIDs), certain antidepressants, chemotherapeutic agents, bronchodilators, antacids, and laxatives. Occlusive or nonocclusive ischemic colitis typically occurs in persons aged >50 years; often presents as acute lower abdominal pain preceding watery, then bloody diarrhea; and generally results in acute inflammatory changes in the sigmoid or left colon while sparing the rectum. Acute diarrhea may accompany colonic diverticulitis and graft-versus-host disease. Acute diarrhea, often associated with systemic compromise, can follow ingestion of toxins including organophosphate insecticides, amanita and other mushrooms, arsenic, and preformed toxins in seafood such as ciguatera (from algae that the fish eat) and scombroid (an excess of histamine due to inadequate refrigeration). Acute anaphylaxis to food ingestion can have a similar presentation. Conditions causing chronic diarrhea can also be confused with acute diarrhea early in their course. This confusion may occur with inflammatory bowel disease (IBD) and some of the other inflammatory chronic diarrheas that may have an abrupt rather than insidious onset and exhibit features that mimic infection.

APPROACH TO THE PATIENT

Acute Diarrhea

The decision to evaluate acute diarrhea depends on its severity and duration and on various host factors (Fig. 46-3). Most episodes of acute diarrhea are mild and self-limited and do not justify the cost and potential morbidity rate of diagnostic or pharmacologic interventions. Indications for evaluation include profuse diarrhea with dehydration, grossly bloody stools, fever $\geq 38.5^{\circ}\text{C}$ ($\geq 101^{\circ}\text{F}$), duration >48 h without improvement, recent antibiotic use, new community outbreaks, associated severe abdominal pain in patients aged >50 years, and elderly (≥ 70 years) or immunocompromised patients. In some cases of moderately severe febrile diarrhea associated with fecal leukocytes (or increased fecal levels of the leukocyte proteins, such as calprotectin) or with gross blood, a diagnostic evaluation might be avoided in favor of an empirical antibiotic trial (see below).

The cornerstone of diagnosis in those suspected of severe acute infectious diarrhea is microbiologic analysis of the stool. Workup

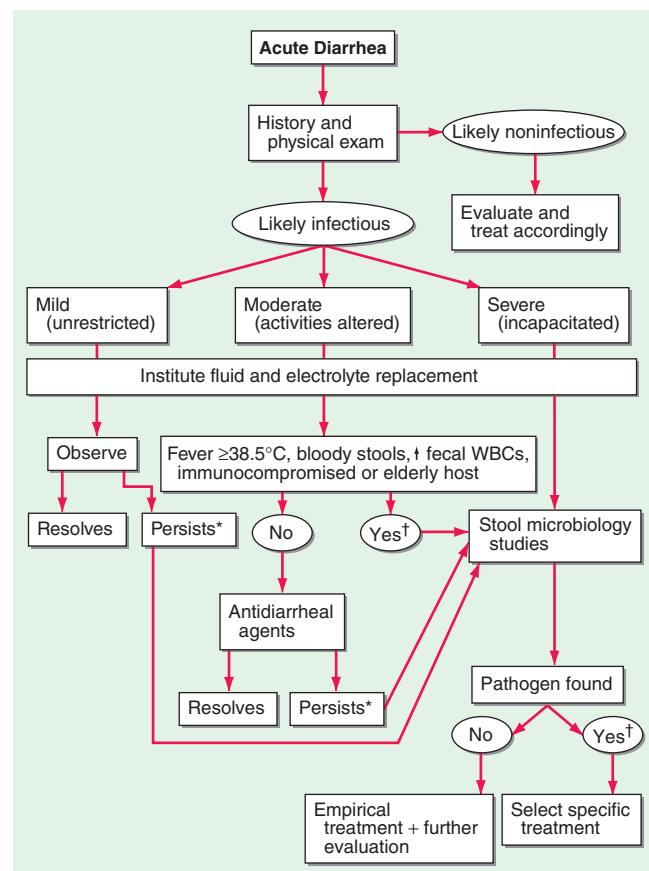


FIGURE 46-3 Algorithm for the management of acute diarrhea. Consider empirical treatment before evaluation with (*) metronidazole and with (†) quinolone. WBCs, white blood cells.

includes cultures for bacterial and viral pathogens; direct inspection for ova and parasites; and immunoassays for certain bacterial toxins (*C. difficile*), viral antigens (rotavirus), and protozoal antigens (*Giardia*, *E. histolytica*). The aforementioned clinical and epidemiologic associations may assist in focusing the evaluation. If a particular pathogen or set of possible pathogens is so implicated, either the whole panel of routine studies may not be necessary or, in some instances, special cultures may be appropriate, as for enterohemorrhagic and other types of *E. coli*, *Vibrio* species, and *Yersinia*. Molecular diagnosis of pathogens in stool can be made by identification of unique DNA sequences, and evolving microarray technologies have led to more rapid, sensitive, specific, and cost-effective diagnosis.

Persistent diarrhea is commonly due to *Giardia* (Chap. 223), but additional causative organisms that should be considered include *C. difficile* (especially if antibiotics had been administered), *E. histolytica*, *Cryptosporidium*, *Campylobacter*, and others. If stool studies are unrevealing, flexible sigmoidoscopy with biopsies and upper endoscopy with duodenal aspirates and biopsies may be indicated. Brainerd diarrhea is an increasingly recognized entity characterized by an abrupt-onset diarrhea that persists for at least 4 weeks, but may last 1–3 years, and is thought to be of infectious origin. It may be associated with subtle inflammation of the distal small intestine or proximal colon.

Structural examination by sigmoidoscopy, colonoscopy, or abdominal computed tomography (CT) scanning (or other imaging approaches) may be appropriate in patients with uncharacterized persistent diarrhea to exclude IBD or as an initial approach in patients with suspected noninfectious acute diarrhea such as might be caused by ischemic colitis, diverticulitis, or partial bowel obstruction.

TREATMENT

Acute Diarrhea

Fluid and electrolyte replacement are of central importance to all forms of acute diarrhea. Fluid replacement alone may suffice for mild cases. Oral sugar-electrolyte solutions (iso-osmolar sport drinks or designed formulations) should be instituted promptly with severe diarrhea to limit dehydration, which is the major cause of death. Profoundly dehydrated patients, especially infants and the elderly, require IV rehydration.

In moderately severe nonfebrile and nonbloody diarrhea, antimotility and antisecretory agents such as loperamide can be useful adjuncts to control symptoms. Such agents should be avoided with febrile dysentery, which may be prolonged by them, and should be used with caution with drugs that increase levels due to cardiotoxicity. Bismuth subsalicylate may reduce symptoms of vomiting and diarrhea but should not be used to treat immunocompromised patients or those with renal impairment because of the risk of bismuth encephalopathy.

Judicious use of antibiotics is appropriate in selected instances of acute diarrhea and may reduce its severity and duration (Fig. 46-3). Many physicians treat moderately to severely ill patients with febrile dysentery empirically without diagnostic evaluation using a quinolone, such as ciprofloxacin (500 mg bid for 3–5 d). Empirical treatment can also be considered for suspected giardiasis with metronidazole (250 mg qid for 7 d). Selection of antibiotics and dosage regimens are otherwise dictated by specific pathogens, geographic patterns of resistance, and conditions found ([Chaps. 133, 161, and 165–171](#)). Because of resistance to first-line treatments, newer agents such as nitazoxanide may be required for *Giardia* and *Cryptosporidium* infections. Antibiotic coverage is indicated, whether or not a causative organism is discovered, in patients who are immunocompromised, have mechanical heart valves or recent vascular grafts, or are elderly. Bismuth subsalicylate may reduce the frequency of traveler's diarrhea. Antibiotic prophylaxis is only indicated for certain patients traveling to high-risk countries in whom the likelihood or seriousness of acquired diarrhea would be especially high, including those with immunocompromise, IBD, hemochromatosis, or gastric achlorhydria. Use of ciprofloxacin, azithromycin, or rifaximin may reduce bacterial diarrhea in such travelers by 90%, though rifaximin is not suitable for invasive disease but rather as treatment for uncomplicated traveler's diarrhea. There is little role for endoscopic evaluation in most circumstances except in immunocompromised patients. Finally, physicians should be vigilant to identify if an outbreak of diarrheal illness is occurring and to alert the public health authorities promptly. This may reduce the ultimate size of the affected population.

CHRONIC DIARRHEA

Diarrhea lasting >4 weeks warrants evaluation to exclude serious underlying pathology. In contrast to acute diarrhea, most of the causes of chronic diarrhea are noninfectious. The classification of chronic diarrhea by pathophysiologic mechanism facilitates a rational approach to management, although many diseases cause diarrhea by more than one mechanism ([Table 46-3](#)).

Secretory Causes Secretory diarrheas are due to derangements in fluid and electrolyte transport across the enterocolonic mucosa. They are characterized clinically by watery, large-volume fecal outputs that are typically painless and persist with fasting. Because there is no malabsorbed solute, stool osmolality is accounted for by normal endogenous electrolytes with no fecal osmotic gap.

MEDICATIONS Side effects from regular ingestion of drugs and toxins are the most common secretory causes of chronic diarrhea. Hundreds of prescription and over-the-counter medications (see earlier section, "Acute Diarrhea, Other Causes") may produce diarrhea. Surreptitious or habitual use of stimulant laxatives (e.g., senna, cascara, bisacodyl,

TABLE 46-3 Major Causes of Chronic Diarrhea According to Predominant Pathophysiologic Mechanism

Secretory Causes

- Exogenous stimulant laxatives
- Chronic ethanol ingestion
- Other drugs and toxins
- Endogenous laxatives (dihydroxy bile acids)
- Idiopathic secretory diarrhea or bile acid diarrhea
- Certain bacterial infections
- Bowel resection, disease, or fistula (↓ absorption)
- Partial bowel obstruction or fecal impaction
- Hormone-producing tumors (carcinoid, VIPoma, medullary cancer of thyroid, mastocytosis, gastrinoma, colorectal villous adenoma)
- Addison's disease
- Congenital electrolyte absorption defects

Osmotic Causes

- Osmotic laxatives (Mg^{2+} , PO_4^{-3} , SO_4^{-2})
- Lactase and other disaccharide deficiencies
- Nonabsorbable carbohydrates (sorbitol, lactulose, polyethylene glycol)
- Gluten and FODMAP intolerance

Steatorrheal Causes

- Intraluminal maldigestion (pancreatic exocrine insufficiency, bacterial overgrowth, bariatric surgery, liver disease)
- Mucosal malabsorption (celiac sprue, Whipple's disease, infections, abetalipoproteinemia, ischemia, drug-induced enteropathy)
- Postmucosal obstruction (1° or 2° lymphatic obstruction)

Inflammatory Causes

- Idiopathic inflammatory bowel disease (Crohn's, chronic ulcerative colitis)
- Lymphocytic and collagenous colitis
- Immune-related mucosal disease (1° or 2° immunodeficiencies, food allergy, eosinophilic gastroenteritis, graft-versus-host disease)
- Infections (invasive bacteria, viruses, and parasites, Brainerd diarrhea)
- Radiation injury
- Gastrointestinal malignancies

Dysmotile Causes

- Irritable bowel syndrome (including postinfectious IBS)
- Visceral neuromyopathies
- Hyperthyroidism
- Drugs (prokinetic agents)
- Postvagotomy

Factitial Causes

- Munchausen
- Eating disorders

Iatrogenic Causes

- Cholecystectomy
- Ileal resection
- Bariatric surgery
- Vagotomy, fundoplication

Abbreviations: FODMAP, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; IBS, irritable bowel syndrome.

ricinoleic acid [castor oil]) must also be considered. Chronic ethanol consumption may cause a secretory-type diarrhea due to enterocyte injury with impaired sodium and water absorption as well as rapid transit and other alterations. Inadvertent ingestion of certain environmental toxins (e.g., arsenic) may lead to chronic rather than acute forms of diarrhea. Certain bacterial infections may occasionally persist and be associated with a secretory-type diarrhea. The oral angiotensin receptor blocker olmesartan is associated with diarrhea due to sprue-like enteropathy.

BOWEL RESECTION, MUCOSAL DISEASE, OR ENTEROCOLIC FISTULA These conditions may result in a secretory-type diarrhea because of inadequate surface for reabsorption of secreted fluids and electrolytes. Unlike other secretory diarrheas, this subset of conditions tends to worsen with eating. With disease (e.g., Crohn's ileitis) or resection of <100 cm of terminal ileum, dihydroxy bile acids may escape absorption and stimulate colonic secretion (cholerheic diarrhea). This mechanism may contribute to so-called *idiopathic secretory diarrhea or bile acid diarrhea (BAD)*, in which bile acids are functionally malabsorbed from a normal-appearing terminal ileum. This *idiopathic bile acid malabsorption (BAM)* may account for an average of 40% of unexplained chronic diarrhea. Reduced negative feedback regulation of bile acid synthesis in hepatocytes by fibroblast growth factor 19 (FGF-19) produced by ileal enterocytes results in a degree of bile-acid synthesis that exceeds the normal capacity for ileal reabsorption, producing BAD. An alternative cause of BAD is a genetic variation in the receptor proteins (β -klotho and fibroblast growth factor 4) on the hepatocyte that normally mediate the effect of FGF-19. Dysfunction of these proteins prevents FGF-19 inhibition of hepatocyte bile acid synthesis. Another mechanism is based on genetic variation in the bile acid receptor (TGR5) in the colon, resulting in accelerated colonic transit.

Partial bowel obstruction, ostomy stricture, or fecal impaction may paradoxically lead to increased fecal output due to fluid hypersecretion.

HORMONES Although uncommon, the classic examples of secretory diarrhea are those mediated by hormones. *Metastatic gastrointestinal carcinoid tumors* or, rarely, *primary bronchial carcinoids* may produce watery diarrhea alone or as part of the carcinoid syndrome that comprises episodic flushing, wheezing, dyspnea, and right-sided valvular heart disease. Diarrhea is due to the release into the circulation of potent intestinal secretagogues including serotonin, histamine, prostaglandins, and various kinins. Pellagra-like skin lesions may rarely occur as the result of serotonin overproduction with niacin depletion. *Gastrinoma*, one of the most common neuroendocrine tumors, most typically presents with refractory peptic ulcers, but diarrhea occurs in up to one-third of cases and may be the only clinical manifestation in 10%. While other secretagogues released with gastrin may play a role, the diarrhea most often results from fat maldigestion owing to pancreatic enzyme inactivation by low intraduodenal pH. The watery diarrhea hypokalemia achlorhydria syndrome, also called *pancreatic cholera*, is due to a non- β cell pancreatic adenoma, referred to as a *VIPoma*, that secretes VIP and a host of other peptide hormones including pancreatic polypeptide, secretin, gastrin, gastrin-inhibitory polypeptide (also called glucose-dependent insulinotropic peptide), neuropeptid Y, calcitonin, and prostaglandins. The secretory diarrhea is often massive with stool volumes >3 L/d; daily volumes as high as 20 L have been reported. Life-threatening dehydration; neuromuscular dysfunction from associated hypokalemia, hypomagnesemia, or hypercalcemia; flushing; and hyperglycemia may accompany a VIPoma. *Medullary carcinoma of the thyroid* may present with watery diarrhea caused by calcitonin, other secretory peptides, or prostaglandins. Prominent diarrhea is often associated with metastatic disease and poor prognosis. *Systemic mastocytosis*, which may be associated with the skin lesion urticaria pigmentosa, may cause diarrhea that is either secretory and mediated by histamine or inflammatory due to intestinal infiltration by mast cells. Large *colorectal villous adenomas* may rarely be associated with a secretory diarrhea that may cause hypokalemia, can be inhibited by NSAIDs, and are apparently mediated by prostaglandins.

CONGENITAL DEFECTS IN ION ABSORPTION Rarely, defects in specific carriers associated with ion absorption cause watery diarrhea from birth. These disorders include defective $\text{Cl}^-/\text{HCO}_3^-$ exchange (*congenital chloridorrhea*) with alkalosis (which results from a mutated *DRA* [down-regulated in adenoma] gene) and defective Na^+/H^+ exchange (*congenital sodium diarrhea*), which results from a mutation in the *NHE3* (sodium-hydrogen exchanger) gene and results in acidosis.

Some hormone deficiencies may be associated with watery diarrhea, such as occurs with adrenocortical insufficiency (Addison's disease) that may be accompanied by skin hyperpigmentation.

Osmotic Causes Osmotic diarrhea occurs when ingested, poorly absorbable, osmotically active solutes draw enough fluid into the lumen to exceed the reabsorptive capacity of the colon. Fecal water output increases in proportion to such a solute load. Osmotic diarrhea characteristically ceases with fasting or with discontinuation of the causative agent.

OSMOTIC LAXATIVES Ingestion of magnesium-containing antacids, health supplements, or laxatives may induce osmotic diarrhea typified by a stool osmotic gap (>50 mosmol/L): serum osmolarity (typically 290 mosmol/kg) – (2 × [fecal sodium + potassium concentration]). Measurement of fecal osmolarity is no longer recommended because, even when measured immediately after evacuation, it may be erroneous because carbohydrates are metabolized by colonic bacteria, causing an increase in osmolarity.

CARBOHYDRATE MALABSORPTION Carbohydrate malabsorption due to acquired or congenital defects in brush-border disaccharidases and other enzymes leads to osmotic diarrhea with a low pH. One of the most common causes of chronic diarrhea in adults is *lactase deficiency*, which affects three-fourths of nonwhites worldwide and 5–30% of persons in the United States; the total lactose load at any one time influences the symptoms experienced. Most patients learn to avoid milk products without requiring treatment with enzyme supplements. Some sugars, such as sorbitol, lactulose, or fructose, are frequently malabsorbed, and diarrhea ensues with ingestion of medications, gum, or candies sweetened with these poorly or incompletely absorbed sugars.

WHEAT AND FODMAP INTOLERANCE Chronic diarrhea, bloating, and abdominal pain are recognized as symptoms of nonceliac gluten intolerance (which is associated with impaired intestinal or colonic barrier function) and intolerance of fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs). The latter's effects represent the interaction between the GI microbiome and the nutrients.

Steatorrheal Causes Fat malabsorption may lead to greasy, foul-smelling, difficult-to-flush diarrhea often associated with weight loss and nutritional deficiencies due to concomitant malabsorption of amino acids and vitamins. Increased fecal output is caused by the osmotic effects of fatty acids, especially after bacterial hydroxylation, and, to a lesser extent, by the neutral fat. Quantitatively, steatorrhea is defined as stool fat exceeding the normal 7 g/d; rapid-transit diarrhea may result in fecal fat up to 14 g/d; daily fecal fat averages 15–25 g with small-intestinal diseases and is often >32 g with pancreatic exocrine insufficiency. Intraluminal maldigestion, mucosal malabsorption, or lymphatic obstruction may produce steatorrhea.

INTRALUMINAL MALDIGESTION This condition most commonly results from pancreatic exocrine insufficiency, which occurs when >90% of pancreatic secretory function is lost. *Chronic pancreatitis*, usually a sequel of ethanol abuse, most frequently causes pancreatic insufficiency. Other causes include *cystic fibrosis*, *pancreatic duct obstruction*, and, rarely, *somatostatinoma*. Bacterial overgrowth in the small intestine may deconjugate bile acids and alter micelle formation, impairing fat digestion; it occurs with stasis from a blind-loop, small-bowel diverticulum or dysmotility and is especially likely in the elderly. Finally, cirrhosis or biliary obstruction may lead to mild steatorrhea due to deficient intraluminal bile acid concentration.

MUCOSAL MALABSORPTION Mucosal malabsorption occurs from a variety of enteropathies, but it most commonly occurs from *celiac disease*. This gluten-sensitive enteropathy affects all ages and is characterized by villous atrophy and crypt hyperplasia in the proximal small bowel and can present with fatty diarrhea associated with multiple nutritional deficiencies of varying severity. Celiac disease is much more frequent than previously thought; it affects ~1% of the population, frequently presents without steatorrhea, can mimic IBS, and has many other GI and extraintestinal manifestations. *Tropical sprue* may produce a similar histologic and clinical syndrome but occurs in residents of or travelers to tropical climates; abrupt onset and response to antibiotics suggest an infectious etiology. *Whipple's disease*, due to

the bacillus *Tropheryma whipplei* and histiocytic infiltration of the small-bowel mucosa, is a less common cause of steatorrhea that most typically occurs in young or middle-aged men; it is frequently associated with arthralgias, fever, lymphadenopathy, and extreme fatigue, and it may affect the CNS and endocardium. A similar clinical and histologic picture results from *Mycobacterium avium-intracellulare* infection in patients with AIDS. *Abetalipoproteinemia* is a rare defect of chylomicron formation and fat malabsorption in children, associated with acanthocytic erythrocytes, ataxia, and retinitis pigmentosa. Several other conditions may cause mucosal malabsorption including infections, especially with protozoa such as *Giardia*, numerous medications (e.g., olmesartan, mycophenolate mofetil, colchicine, cholestyramine, neomycin), idiopathic enteropathies, amyloidosis, and chronic ischemia.

POSTMUCOSAL LYMPHATIC OBSTRUCTION The pathophysiology of this condition, which is due to the rare *congenital intestinal lymphangiectasia* or to *acquired lymphatic obstruction* secondary to trauma, tumor, cardiac disease, or infection, leads to the unique constellation of fat malabsorption with enteric losses of protein (often causing edema) and lymphocytopenia. Carbohydrate and amino acid absorption are preserved.

Inflammatory Causes Inflammatory diarrheas are generally accompanied by pain, fever, bleeding, or other manifestations of inflammation. The mechanism of diarrhea may not only be exudation but, depending on lesion site, may include fat malabsorption, disrupted fluid/electrolyte absorption, and hypersecretion or hypermotility from release of cytokines and other inflammatory mediators. The unifying feature on stool analysis is the presence of leukocytes or leukocyte-derived proteins such as calprotectin. With severe inflammation, exudative protein loss can lead to anasarca (generalized edema). Any middle-aged or older person with chronic inflammatory-type diarrhea, especially with blood, should be carefully evaluated to exclude a colorectal tumor.

IDIOPATHIC INFLAMMATORY BOWEL DISEASE The illnesses in this category, which include *Crohn's disease* and *chronic ulcerative colitis*, are among the most common organic causes of chronic diarrhea in adults and range in severity from mild to fulminant and life-threatening. They may be associated with uveitis, polyarthralgias, cholestatic liver disease (primary sclerosing cholangitis), and skin lesions (erythema nodosum, pyoderma gangrenosum). *Microscopic colitis*, including both lymphocytic and *collagenous colitis*, is an increasingly recognized cause of chronic watery diarrhea, especially in middle-aged women and those on NSAIDs, statins, proton pump inhibitors (PPIs), and selective serotonin reuptake inhibitors (SSRIs); biopsy of a normal-appearing colon is required for histologic diagnosis. It may coexist with symptoms suggesting IBS or with celiac sprue or drug-induced enteropathy. It typically responds well to anti-inflammatory drugs (e.g., bismuth), the opioid agonist loperamide, or budesonide.

PRIMARY OR SECONDARY FORMS OF IMMUNODEFICIENCY Immunodeficiency may lead to prolonged infectious diarrhea. With selective IgA deficiency or common variable *hypogammaglobulinemia*, diarrhea is particularly prevalent and often the result of giardiasis, bacterial overgrowth, or sprue.

EOSINOPHILIC GASTROENTERITIS Eosinophil infiltration of the mucosa, muscularis, or serosa at any level of the GI tract may cause diarrhea, pain, vomiting, or ascites. Affected patients often have an atopic history, Charcot-Leyden crystals due to extruded eosinophil contents may be seen on microscopic inspection of stool, and peripheral eosinophilia is present in 50–75% of patients. While hypersensitivity to certain foods occurs in adults, true food allergy causing chronic diarrhea is rare.

OTHER CAUSES Chronic inflammatory diarrhea may be caused by *radiation enterocolitis*, *chronic graft-versus-host disease*, autoimmune

or idiopathic enteropathies, *Behcet's syndrome*, and *Cronkhite-Canada syndrome*, among others.

Dysmotility Causes Rapid transit may accompany many diarrheas as a secondary or contributing phenomenon, but primary dysmotility is an unusual etiology of true diarrhea. Stool features often suggest a secretory diarrhea, but mild steatorrhea of up to 14 g of fat per day can be produced by maldigestion from rapid transit alone. *Hyperthyroidism*, *carcinoid syndrome*, and certain drugs (e.g., prostaglandins, prokinetic agents) may produce hypermotility with resultant diarrhea. Primary visceral neuromyopathies or idiopathic acquired intestinal pseudoobstruction may lead to stasis with secondary bacterial overgrowth causing diarrhea. *Diabetic diarrhea*, often accompanied by peripheral and generalized autonomic neuropathies, may occur in part because of intestinal dysmotility.

The exceedingly common IBS (10% point prevalence, 1–2% per year incidence) is characterized by disturbed intestinal and colonic motor and sensory responses to various stimuli. Symptoms of stool frequency typically cease at night, alternate with periods of constipation, are accompanied by abdominal pain relieved with defecation, and rarely result in weight loss.

Factitious Causes Factitial diarrhea accounts for up to 15% of unexplained diarrheas referred to tertiary care centers. Either as a form of *Munchausen syndrome* (deception or self-injury for secondary gain) or *eating disorders*, some patients covertly self-administer laxatives alone or in combination with other medications (e.g., diuretics) or surreptitiously add water or urine to stool sent for analysis. Such patients are typically women, often with histories of psychiatric illness, and disproportionately from careers in health care. Hypotension and hypokalemia are common co-presenting features. The evaluation of such patients may be difficult: contamination of the stool with water or urine is suggested by very low or high stool osmolarity, respectively. Such patients often deny this possibility when confronted, but they do benefit from psychiatric counseling when they acknowledge their behavior.

APPROACH TO THE PATIENT

CHRONIC DIARRHEA

The laboratory tools available to evaluate the very common problem of chronic diarrhea are extensive, and many are costly and invasive. As such, the diagnostic evaluation must be rationally directed by a careful history, including medications, and physical examination (Fig. 46-4). When this strategy is unrevealing, simple triage tests are often warranted to direct the choice of more complex investigations (Fig. 46-4). The history, physical examination (Table 46-4), and routine blood studies should attempt to characterize the mechanism of diarrhea, identify diagnostically helpful associations, and assess the patient's fluid/electrolyte and nutritional status. Patients should be questioned about the onset, duration, pattern, aggravating (especially diet) and relieving factors, and stool characteristics of their diarrhea. The presence or absence of fecal incontinence, fever, weight loss, pain, certain exposures (travel, medications, contacts with diarrhea), and common extraintestinal manifestations (skin changes, arthralgias, oral aphthous ulcers) should be noted. A family history of IBD or celiac disease may indicate those possibilities. Physical findings may offer clues such as a thyroid mass, wheezing, heart murmurs, edema, hepatomegaly, abdominal masses, lymphadenopathy, mucocutaneous abnormalities, perianal fistulas, or anal sphincter laxity. Peripheral blood leukocytosis, elevated sedimentation rate, or C-reactive protein suggests inflammation; anemia reflects blood loss or nutritional deficiencies; or eosinophilia may occur with parasitoses, neoplasia, collagen-vascular disease, allergy, or eosinophilic gastroenteritis. Blood chemistries may demonstrate electrolyte, hepatic, or other metabolic disturbances. Measuring IgA tissue transglutaminase antibodies

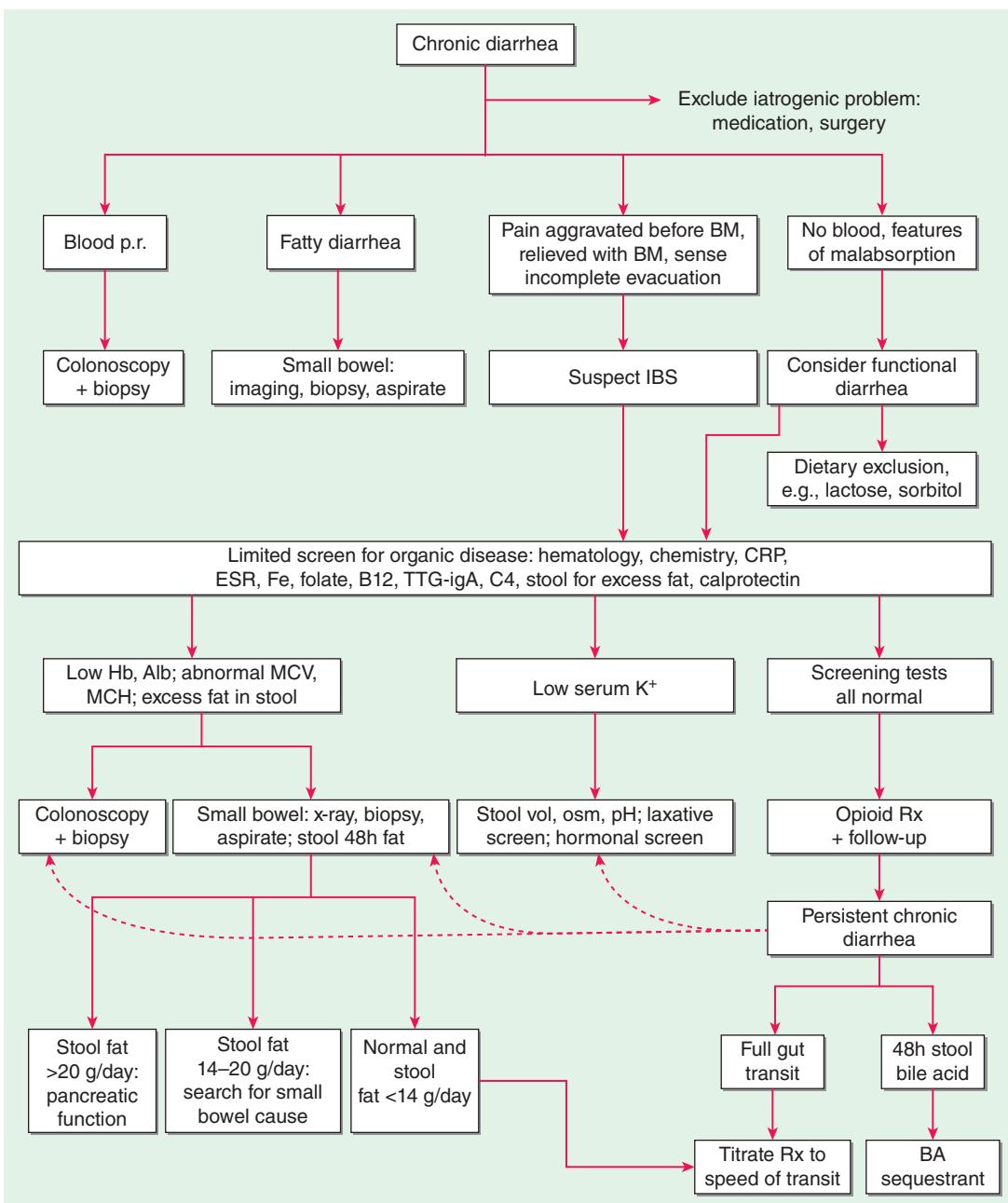


FIGURE 46-4 Algorithm for management of chronic diarrhea. Patients undergo an initial evaluation based on different symptom presentations, leading to selection of patients for imaging, biopsy analysis, and limited screens for organic diseases. Alb, albumin; BA, bile acid; BM, bowel movement; C4, 7 α -hydroxy-4-cholesten-3-one; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; Hb, hemoglobin; Hx, history; IBS, irritable bowel syndrome; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume; osm, osmolality; p.r., per rectum; Rx, treatment; TTG, tissue transglutaminase. (Reproduced with permission from M Camilleri, JH Sellin, KE Barrett: Pathophysiology, evaluation, and management of chronic watery diarrhea. *Gastroenterology* 152:515, 2017.)

TABLE 46-4 Physical Examination in Patients with Chronic Diarrhea

- Are there general features to suggest malabsorption or inflammatory bowel disease (IBD) such as anemia, dermatitis herpetiformis, edema, or clubbing?
- Are there features to suggest underlying autonomic neuropathy or collagen-vascular disease in the pupils, orthostasis, skin, hands, or joints?
- Is there an abdominal mass or tenderness?
- Are there any abnormalities of rectal mucosa, rectal defects, or altered anal sphincter functions?
- Are there any mucocutaneous manifestations of systemic disease such as dermatitis herpetiformis (celiac disease), erythema nodosum (ulcerative colitis), flushing (carcinoid), or oral ulcers for IBD or celiac disease?

may help detect celiac disease. Bile acid diarrhea is confirmed by a scintigraphic radiolabeled bile acid retention test; however, this is not available in many countries. Alternative approaches are a screening blood test (serum C4 or FGF-19), measurement of fecal bile acids, or a therapeutic trial with a bile acid sequestrant (e.g., cholestyramine, colestipol or colestevam).

A therapeutic trial is often appropriate, definitive, and highly cost-effective when a specific diagnosis is suggested on the initial physician encounter. For example, chronic watery diarrhea, which ceases with fasting in an otherwise healthy young adult, may justify a trial of a lactose-restricted diet; bloating and diarrhea persisting since a mountain backpacking trip may warrant a trial of metronidazole for likely giardiasis; and postprandial diarrhea persisting

following resection of terminal ileum might be due to bile acid malabsorption and be treated with cholestyramine, colestipol, or colesevelam before further evaluation. Persistent symptoms require additional investigation.

Certain diagnoses may be suggested on the initial encounter (e.g., idiopathic IBD); however, additional focused evaluations may be necessary to confirm the diagnosis and characterize the severity or extent of disease so that treatment can be best guided. Patients suspected of having IBS should be initially evaluated with flexible sigmoidoscopy with colorectal biopsies to exclude IBD, or particularly microscopic colitis, which is clinically indistinguishable from IBS with diarrhea or functional diarrhea; those with normal findings might be reassured and, as indicated, treated empirically with antispasmodics, antidiarrheals, or antidepressants (e.g., tricyclic agents). Any patient who presents with chronic diarrhea and hematochezia should be evaluated with stool microbiologic studies and colonoscopy.

In an estimated two-thirds of cases, the cause for chronic diarrhea remains unclear after the initial encounter, and further testing is required. Quantitative stool collection and analyses can yield important objective data that may establish a diagnosis or characterize the type of diarrhea as a triage for focused additional studies (Fig. 46-4). If stool weight is >200 g/d, additional stool analyses should be performed that might include electrolyte concentration, pH, occult blood testing, leukocyte inspection (or leukocyte protein assay), fat quantitation, and laxative screens.

For secretory diarrheas (watery, normal osmotic gap), possible medication-related side effects or surreptitious laxative use should be reconsidered. Microbiologic studies should be done including fecal bacterial cultures (including media for *Aeromonas* and *Plesiomonas*), inspection for ova and parasites, and *Giardia* antigen assay (the most sensitive test for giardiasis). Small-bowel bacterial overgrowth can be excluded by intestinal aspirates with quantitative cultures or with glucose or lactulose breath tests involving measurement of breath hydrogen, methane, or other metabolite. However, interpretation of these breath tests may be confounded by disturbances of intestinal transit. Upper endoscopy and colonoscopy with biopsies and small-bowel x-rays (formerly barium, but increasingly CT with enterography or magnetic resonance with enteroclysis) are helpful to rule out structural or occult inflammatory disease. When suggested by history or other findings, screens for peptide hormones should be pursued (e.g., serum gastrin, VIP, calcitonin, thyroid hormone/thyroid-stimulating hormone, urinary 5-hydroxyindolacetic acid, histamine).

Further evaluation of osmotic diarrhea should include tests for lactose intolerance and magnesium ingestion, the two most common causes. Low fecal pH suggests carbohydrate malabsorption; lactose malabsorption can be confirmed by lactose breath testing or by a therapeutic trial with lactose exclusion and observation of the effect of lactose challenge (e.g., a liter of milk). Lactase determination on small-bowel biopsy is not generally available. If fecal magnesium or laxative levels are elevated, inadvertent or surreptitious ingestion should be considered and psychiatric help should be sought.

For those with proven fatty diarrhea, endoscopy with small-bowel biopsy (including aspiration for quantitative cultures, if available) should be performed; if this procedure is unrevealing, a small-bowel radiograph is often an appropriate next step. If small-bowel studies are negative or if pancreatic disease is suspected, pancreatic exocrine insufficiency should be excluded with direct tests, such as the secretin-cholecystokinin stimulation test or a variation that could be performed endoscopically. In general, indirect tests such as assay of fecal elastase or chymotrypsin activity or a bentiromide test have fallen out of favor because of low sensitivity and specificity.

Chronic inflammatory-type diarrheas should be suspected by the presence of blood or leukocytes in the stool. Such findings warrant stool cultures; inspection for ova and parasites; *C. difficile* toxin assay; colonoscopy with biopsies; and, if indicated, small-bowel imaging studies.

TREATMENT

Chronic Diarrhea

Treatment of chronic diarrhea depends on the specific etiology and may be curative, suppressive, or empirical. If the cause can be eradicated, treatment is curative as with resection of a colorectal cancer, antibiotic administration for Whipple's disease or tropical sprue, or discontinuation of a drug. For many chronic conditions, diarrhea can be controlled by suppression of the underlying mechanism. Examples include elimination of dietary lactose for lactase deficiency or gluten for celiac sprue, use of glucocorticoids or other anti-inflammatory agents for idiopathic IBDs, bile acid sequestrants for bile acid malabsorption, PPIs for the gastric hypersecretion of gastrinomas, somatostatin analogues such as octreotide for malignant carcinoid syndrome, prostaglandin inhibitors such as indomethacin for medullary carcinoma of the thyroid, and pancreatic enzyme replacement for pancreatic insufficiency. When the specific cause or mechanism of chronic diarrhea evades diagnosis, empirical therapy may be beneficial. Mild opiates, such as diphenoxylate or loperamide, are often helpful in mild or moderate watery diarrhea. For those with more severe diarrhea, codeine or tincture of opium may be beneficial. Such antimotility agents should be avoided with severe IBD, because toxic megacolon may be precipitated. Clonidine, an α_2 -adrenergic agonist, may allow control of diabetic diarrhea, although the medication may be poorly tolerated because it causes postural hypotension. The 5-HT₃ receptor antagonists (e.g., alosetron, ondansetron) may relieve diarrhea and urgency in patients with IBS diarrhea. Other medications approved for the treatment of diarrhea associated with IBS are the nonabsorbed antibiotic, rifaximin, and the mixed μ -opioid receptor (OR) and κ -OR agonist and δ -OR antagonist, eluxadoline. The latter may induce sphincter of Oddi spasm and subsequent acute pancreatitis, usually in patients with prior cholecystectomy. For all patients with chronic diarrhea, fluid and electrolyte repletion is an important component of management (see "Acute Diarrhea," earlier). Replacement of fat-soluble vitamins may also be necessary in patients with chronic steatorrhea.

CONSTIPATION

■ DEFINITION

Constipation is a common complaint in clinical practice and usually refers to persistent, difficult, infrequent, or seemingly incomplete defecation. Because of the wide range of normal bowel habits, constipation is difficult to define precisely. Most persons have at least three bowel movements per week; however, low stool frequency alone is not the sole criterion for the diagnosis of constipation. Many constipated patients have a normal frequency of defecation but complain of excessive straining, hard stools, lower abdominal fullness, or a sense of incomplete evacuation. The individual patient's symptoms must be analyzed in detail to ascertain what is meant by "constipation" or "difficulty" with defecation.

Stool form and consistency are well correlated with the time elapsed from the preceding defecation. Hard, pellet stools occur with slow transit, whereas loose, watery stools are associated with rapid transit. Both small pellet or very large stools are more difficult to expel than normal stools.

The perception of hard stools or excessive straining is more difficult to assess objectively, and the need for enemas or digital disimpaction is a clinically useful way to corroborate the patient's perceptions of difficult defecation.

Psychosocial or cultural factors may also be important. A person whose parents attached great importance to daily defecation will become greatly concerned when he or she misses a daily bowel movement; some children withhold stool to gain attention or because of fear of pain from anal irritation; and some adults habitually ignore or delay the call to have a bowel movement.

■ CAUSES

Pathophysiologically, chronic constipation generally results from inadequate fiber or fluid intake or from disordered colonic transit or anorectal function. These result from neurogastroenterologic disturbance, certain drugs, advancing age, or in association with a large number of systemic diseases that affect the GI tract (Table 46-5). Constipation of recent onset may be a symptom of significant organic disease such as tumor, anorectal irritation, or stricture. In *idiopathic constipation*, a subset of patients exhibits delayed emptying of the ascending and transverse colon with prolongation of transit (often in the proximal colon) and a reduced frequency of propulsive HAPCs. *Outlet obstruction to defecation* (also called *evacuation disorders*) accounts for about a quarter of cases presenting with constipation in tertiary care and may cause delayed colonic transit, which is usually corrected by biofeedback retraining of the disordered defecation. Constipation of any cause may be exacerbated by hospitalization or chronic illnesses that lead to physical or mental impairment and result in inactivity or physical immobility.

APPROACH TO THE PATIENT

Constipation

A careful history should explore the patient's symptoms and confirm whether she or he is indeed constipated based on frequency (e.g., fewer than three bowel movements per week), consistency (lumpy/hard), excessive straining, prolonged defecation time, or need to support the perineum or digitate the anorectum to facilitate stool evacuation. These latter items identified in the history suggest the presence of a rectal evacuation disorder. In the vast majority of cases (probably >90%), there is no underlying cause (e.g., cancer, depression, or hypothyroidism), and constipation responds to ample hydration, exercise, and supplementation of dietary fiber (15–25 g/d). A good diet and medication history and attention to psychosocial issues are key. Physical examination and, particularly, rectal examination are mandatory and should exclude fecal impaction and most of the important diseases that present with constipation and possibly indicate features suggesting an evacuation disorder (e.g., high anal sphincter tone, failure of perineal descent, or paradoxical puborectalis contraction or puborectalis tenderness during straining to simulate stool evacuation).

The presence of weight loss, rectal bleeding, or anemia with constipation mandates either flexible sigmoidoscopy plus

barium enema or colonoscopy alone, particularly in patients aged >40 years, to exclude structural diseases such as cancer or strictures. Colonoscopy alone is most cost-effective in this setting because it provides an opportunity to biopsy mucosal lesions, perform polypectomy, or dilate strictures. Barium enema has advantages over colonoscopy in the patient with isolated constipation because it is less costly and identifies colonic dilation and all significant mucosal lesions or strictures that are likely to present with constipation. Melanosis coli, or pigmentation of the colon mucosa, indicates the use of anthraquinone laxatives such as cascara or senna; however, this is usually apparent from a careful history. An unexpected disorder such as megacolon or cathartic colon may also be detected by colonic radiographs. Measurement of serum calcium, potassium, and thyroid-stimulating hormone levels will identify rare patients with metabolic disorders.

Patients with more troublesome constipation may not respond to fiber alone and may be helped by a bowel-training regimen, which involves taking an osmotic laxative (e.g., magnesium salts, lactulose, sorbitol, polyethylene glycol) and evacuating with enema or suppository (e.g., glycerin or bisacodyl) as needed. After breakfast, a distraction-free 15–20 min on the toilet without straining is encouraged. Excessive straining may lead to development of hemorrhoids and, if there is weakness of the pelvic floor or injury to the pudendal nerve, may result in obstructed defecation from descending perineum syndrome several years later. Those few who do not benefit from the simple measures delineated above or require long-term treatment or fail to respond to potent laxatives should undergo further investigation (Fig. 46-5). Novel agents that induce secretion (e.g., lubiprostone, a chloride channel activator, or linaclotide, a guanylate cyclase C agonist that activates chloride secretion) are also available.

■ INVESTIGATION OF SEVERE CONSTIPATION

A small minority (probably <5%) of patients have severe or “intractable” constipation; about 25% have evacuation disorders. These are the patients most likely to require evaluation by gastroenterologists or in referral centers. Further observation of the patient may occasionally

TABLE 46-5 Causes of Constipation in Adults

TYPES OF CONSTIPATION AND CAUSES	EXAMPLES
Recent Onset	
Colonic obstruction	Neoplasm; stricture: ischemic, diverticular, inflammatory
Anal sphincter spasm	Anal fissure, painful hemorrhoids
Medications	
Chronic	
Irritable bowel syndrome	Constipation-predominant, alternating
Medications	Ca ²⁺ blockers, antidepressants
Colonic pseudoobstruction	Slow-transit constipation, megacolon (rare Hirschsprung's, Chagas' diseases)
Disorders of rectal evacuation	Pelvic floor dysfunction; anismus; descending perineum syndrome; rectal mucosal prolapse; rectocele
Endocrinopathies	Hypothyroidism, hypercalcemia, pregnancy
Psychiatric disorders	Depression, eating disorders, drugs
Neurologic disease	Parkinsonism, multiple sclerosis, spinal cord injury
Generalized muscle disease	Progressive systemic sclerosis

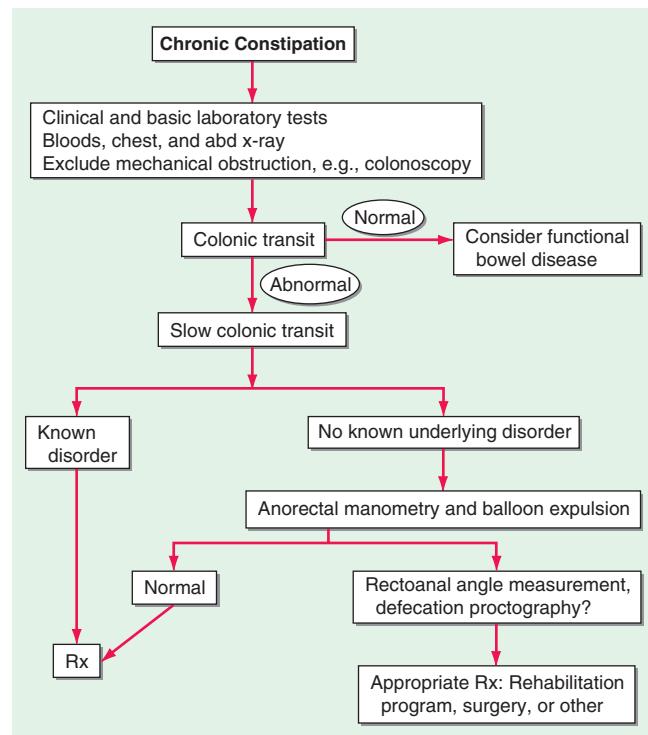


FIGURE 46-5 Algorithm for the management of constipation. abd, abdominal; Rx, treatment.

reveal a previously unrecognized cause, such as an evacuation disorder, laxative abuse, malingering, or psychological disorder. In these patients, evaluations of the physiologic function of the colon and pelvic floor and of psychological status aid in the rational choice of treatment. Even among these highly selected patients with severe constipation, a cause can be identified in only about one-third of tertiary referral patients, with the others being diagnosed with normal transit constipation. Since evacuation disorders also retard colonic transit through the left colon or the entire colon, anorectal and pelvic floor testing should precede transit measurements if there is clinical suspicion of an evacuation disorder. If an evacuation disorder is identified on testing, colonic transit may be unnecessary.

Measurement of Colonic Transit Radiopaque marker transit tests are easy, repeatable, generally safe, inexpensive, reliable, and highly applicable in evaluating constipated patients in clinical practice. Several validated methods are very simple. For example, radiopaque markers are ingested; an abdominal flat film taken 5 days later should indicate passage of 80% of the markers out of the colon without the use of laxatives or enemas. This test does not provide useful information about the transit profile of the stomach and small bowel. An alternative approach involves ingestion of 24 radiopaque markers on 3 successive days and an abdominal radiograph on the fourth day. The number of markers counted in the radiograph is an estimate of the colonic transit in hours. The collection of gas in the rectum between the level of the ischial spines and the lower border of the sacroiliac joints may suggest the presence of a rectal evacuation disorder as the cause of constipation.

Radioscintigraphy with a delayed-release capsule containing radiolabeled particles has been used to noninvasively characterize normal, accelerated, or delayed colonic function over 24–48 h with low radiation exposure. This approach simultaneously assesses gastric, small bowel (which may be important in ~20% of patients with delayed colonic transit because they reflect a more generalized GI motility disorder), and colonic transit. The disadvantages are the greater cost and the need for specific materials prepared in a nuclear medicine laboratory.

Anorectal and Pelvic Floor Tests Pelvic floor dysfunction is suggested by the inability to evacuate the rectum, a feeling of persistent rectal fullness, rectal pain, the need to extract stool from the rectum digitally, application of pressure on the posterior wall of the vagina, support of the perineum during straining, and excessive straining. These significant symptoms should be contrasted with the simple sense of incomplete rectal evacuation, which is common in IBS.

Formal psychological evaluation may identify eating disorders, “control issues,” depression, or posttraumatic stress disorders that may respond to cognitive or other intervention and may be important in restoring quality of life to patients who might present with chronic constipation.

A simple clinical test in the office to document a nonrelaxing puborectalis muscle is to have the patient strain to expel the index finger during a digital rectal examination. Motion of the puborectalis posteriorly during straining indicates proper coordination of the pelvic floor muscles. Motion anteriorly with paradoxical contraction or limited perineal descent (<1.5 cm) during simulated evacuation indicates pelvic floor dysfunction.

Measurement of perineal descent is relatively easy to gauge clinically by placing the patient in the left decubitus position and watching the perineum to detect inadequate descent (<1.5 cm, a sign of pelvic floor dysfunction) or perineal ballooning during straining relative to bony landmarks (>4 cm, suggesting excessive perineal descent).

A useful overall test of evacuation is the balloon expulsion test. A balloon-tipped urinary catheter is placed and inflated with 50 mL of water. Normally, a patient can expel it while seated on a toilet or in the left lateral decubitus position. In the lateral position, the weight needed to facilitate expulsion of the balloon is determined; normally, expulsion occurs with <200 g added or unaided within 1 minute.

Anorectal manometry, when used in the evaluation of patients with severe constipation, may find an excessively high resting (>80 mmHg) or squeeze anal sphincter tone, suggesting anismus (anal

sphincter spasm). This test also identifies rare syndromes, such as adult Hirschsprung’s disease, by the absence of the rectoanal inhibitory reflex.

Defecography (a dynamic barium enema including lateral views obtained during barium expulsion or a magnetic resonance defecogram) reveals “soft abnormalities” in many patients; the most relevant findings are the measured changes in rectoanal angle, anatomic defects of the rectum such as internal mucosal prolapse, and enteroceles or rectoceles. Surgically remediable conditions are identified in only a few patients. These include severe, whole-thickness intussusception with complete outlet obstruction due to funnel-shaped plugging at the anal canal or an extremely large rectocele that fills preferentially during attempts at defecation instead of expulsion of the barium through the anus. In summary, defecography requires an interested and experienced radiologist, and abnormalities are not pathognomonic for pelvic floor dysfunction. The most common cause of outlet obstruction is failure of the puborectalis muscle to relax; this is not identified by barium defecography but can be demonstrated by magnetic resonance defecography, which provides more information about the structure and function of the pelvic floor, distal colorectum, and anal sphincters.

Neurologic testing (electromyography) is more helpful in the evaluation of patients with incontinence than of those with symptoms suggesting obstructed defecation. The absence of neurologic signs in the lower extremities suggests that any documented denervation of the puborectalis results from pelvic (e.g., obstetric) injury or from stretching of the pudendal nerve by chronic, long-standing straining. Constipation is common among patients with spinal cord injuries, neurologic diseases such as Parkinson’s disease, multiple sclerosis, and diabetic neuropathy.

Spinal-evoked responses during electrical rectal stimulation or stimulation of external anal sphincter contraction by applying magnetic stimulation over the lumbosacral cord identify patients with limited sacral neuropathies with sufficient residual nerve conduction to attempt biofeedback training.

In summary, a balloon expulsion test is an important screening test for anorectal dysfunction. Rarely, an anatomic evaluation of the rectum or anal sphincters and an assessment of pelvic floor relaxation are the tools for evaluating patients in whom obstructed defecation is suspected and is associated with symptoms of rectal mucosal prolapse, pressure of the posterior wall of the vagina to facilitate defecation (suggestive of anterior rectocele), or prior pelvic surgery that may be complicated by enterocele.

TREATMENT

Constipation

After the cause of constipation is characterized, a treatment decision can be made. Slow-transit constipation requires aggressive medical or surgical treatment; anismus or pelvic floor dysfunction usually responds to biofeedback management (Fig. 46-5). The remaining ~60% of patients with constipation have normal colonic transit and can be treated symptomatically. Patients with spinal cord injuries or other neurologic disorders require a dedicated bowel regimen that often includes rectal stimulation, enema therapy, and carefully timed laxative therapy.

Patients with constipation are treated with bulk (fiber, psyllium), osmotic (milk of magnesia, lactulose, polyethylene glycol), secretory (lubiprostone, linaclootide, plecanatide, tenapanor), and prokinetic or stimulant laxatives (including diphenyl methanes such as bisacodyl and sodium picosulfate and 5-HT₄ agonists prucalopride and tegaserod). If a 3- to 6-month trial of medical therapies fails, unassociated with obstructed defecation, the patient should be considered for laparoscopic colectomy with ileorectostomy; however, this should not be undertaken for pain or if there is continued evidence of an evacuation disorder or a generalized GI dysmotility. Referral to a specialized center for further tests of colonic motor function is warranted. The decision to resort to surgery is facilitated by the presence of megacolon and megarectum. The complications after surgery include small-bowel obstruction (11%) and fecal

soiling, particularly at night during the first postoperative year. Frequency of defecation is 3–8 per day during the first year, dropping to 1–3 per day from the second year after surgery.

Patients who have a combined (evacuation and transit/motility) disorder should first pursue pelvic floor retraining (biofeedback and muscle relaxation), psychological counseling, and dietetic advice. If symptoms are intractable despite biofeedback and optimized medical therapy, colectomy and ileorectalostomy could be considered as long as the evacuation disorder is resolved and optimized medical therapy is unsuccessful. In patients with pelvic floor dysfunction alone, biofeedback training has a 70–80% success rate, measured by the acquisition of comfortable stool habits. Attempts to manage pelvic floor dysfunction with operations (internal anal sphincter or puborectalis muscle division) or injections with botulinum toxin have achieved only mediocre success and have been largely abandoned.

FURTHER READING

- ASSI R et al: Sexually transmitted infections of the anus and rectum. *World J Gastroenterol* 20:15262, 2014.
- BHARUCHA AE, RAO SS: An update on anorectal disorders for gastroenterologists. *Gastroenterology* 146:37, 2014.
- BHARUCHA AE et al: American Gastroenterological Association technical review on constipation. *Gastroenterology* 144:218, 2013.
- BOECKXSTAENS G et al: Fundamentals of neurogastroenterology: Physiology/motility—sensation. *Gastroenterology* 150:1292, 2016.
- CAMILLERI M et al: Chronic constipation. *Nat Rev Dis Primers* 3:17095, 2017.
- CAMILLERI M et al: Pathophysiology, evaluation, and management of chronic watery diarrhea. *Gastroenterology* 152:515, 2017.
- PEERY AF et al: Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: Update 2018. *Gastroenterology* 156:254, 2019.
- RIDDLE MS et al: ACG Clinical Guideline: Diagnosis, treatment, and prevention of acute diarrheal infections in adults. *Am J Gastroenterol* 111:602, 2016.
- RUBIO-TAPIA A et al: American College of Gastroenterology. ACG clinical guidelines: Diagnosis and management of celiac disease. *Am J Gastroenterol* 108:656, 2013.
- SMALLEY W et al: AGA Clinical Practice Guidelines on the laboratory evaluation of functional diarrhea and diarrhea-predominant irritable bowel syndrome in adults (IBS-D). *Gastroenterology* 157:851, 2019.
- UZZAN M et al: Gastrointestinal disorders associated with common variable immune deficiency (CVID) and chronic granulomatous disease (CGD). *Curr Gastroenterol Rep* 18:17, 2016.

those with known causes, particularly when the source is neoplastic. Weight loss in older persons is associated with a variety of deleterious effects, including falls and fractures, pressure ulcers, impaired immune function, and decreased functional status. Not surprisingly, significant weight loss is associated with increased mortality, which can range from 9% to as high as 38% within 1–2.5 years in the absence of clinical awareness and attention.

PHYSIOLOGY OF WEIGHT REGULATION WITH AGING

(See also Chaps. 401 and 476) Among healthy aging people, total body weight peaks in the sixth decade of life and generally remains stable until the ninth decade, after which it gradually falls. In contrast, lean body mass (fat-free mass) begins to decline at a rate of 0.3 kg per year in the third decade, and the rate of decline increases further beginning at age 60 in men and age 65 in women. These changes in lean body mass largely reflect the age-dependent decline in growth hormone secretion and, consequently, circulating levels of insulin-like growth factor type I (IGF-I) that occur with normal aging. Loss of sex steroids, at menopause in women and more gradually in men, also contributes to these changes in body composition. In the healthy elderly, an increase in fat tissue balances the loss in lean body mass until very old age, when loss of both fat and skeletal muscle occurs. Age-dependent changes also occur at the cellular level. Telomeres shorten, and body cell mass—the fat-free portion of cells—declines steadily with aging.

Between ages 20 and 80, mean energy intake is reduced by up to 1200 kcal/d in men and 800 kcal/d in women. Decreased hunger is a reflection of reduced physical activity and loss of lean body mass, producing lower demand for calories and food intake. Several important age-associated physiologic changes also predispose elderly persons to weight loss, such as declining chemosensory function (smell and taste), reduced efficiency of chewing, slowed gastric emptying, and alterations in the neuroendocrine axis, including changes in levels of leptin, cholecystokinin, neuropeptide Y, and other hormones and peptides. These changes are associated with early satiety and a decline in both appetite and the hedonistic appreciation of food. Collectively, they contribute to the “anorexia of aging.” As noted below, these physiologic changes with aging may be accompanied by social isolation, poverty, and immobility, further contributing to undernutrition.

CAUSES OF UNINTENTIONAL WEIGHT LOSS

Most causes of UWL belong to one of four categories: (1) malignant neoplasms, (2) chronic inflammatory or infectious diseases, (3) metabolic disorders (e.g., hyperthyroidism and diabetes), or (4) psychiatric disorders (Table 47-1). Not infrequently, more than one of these causes can be responsible for UWL. Depending upon patient populations, UWL is caused by malignant disease in a quarter of patients and by organic disease in one-third, with the remainder due to psychiatric disease, medications, or uncertain causes. Risk factors for undiagnosed cancer include a history of smoking, particularly for men, localizing symptoms, and abnormal laboratory tests.

The most common malignant causes of UWL are gastrointestinal, hepatobiliary, hematologic, lung, breast, genitourinary, ovarian, and prostate. Half of all patients with cancer lose some body weight; one-third lose more than 5% of their original body weight, and up to 20% of all cancer deaths are caused directly by cachexia (through immobility and/or cardiac/respiratory failure). The greatest incidence of weight loss is seen among patients with solid tumors. Malignancy that reveals itself through significant weight loss usually has a very poor prognosis.

In addition to malignancies, gastrointestinal diseases are among the most prominent causes of UWL. Peptic ulcer disease, inflammatory bowel disease, dysmotility syndromes, chronic pancreatitis, celiac disease, constipation, and atrophic gastritis are some of the more common entities. Oral and dental problems are easily overlooked and may manifest with halitosis, poor oral hygiene, xerostomia, inability to chew, reduced masticatory force, nonocclusion, temporomandibular joint syndrome, edentulousness, and pain due to caries or abscesses.

Tuberculosis, fungal diseases, parasites, subacute bacterial endocarditis, and HIV are well-documented causes of UWL. Cardiovascular

47

Unintentional Weight Loss

J. Larry Jameson

Involuntary or unintentional weight loss (UWL) is frequently insidious and can have important implications, often serving as a harbinger of serious underlying disease. Clinically important weight loss is defined as the loss of 10 pounds (4.5 kg) or >5% of one's body weight over a period of 6–12 months. UWL is encountered in up to 8% of all adult outpatients and 27% of frail persons aged ≥65 years. There is no identifiable cause in up to one-quarter of patients despite extensive investigation. Conversely, up to half of people who claim to have lost weight have no documented evidence of weight loss. People with no known cause of weight loss generally have a better prognosis than do

TABLE 47-1 Causes of Involuntary Weight Loss

Cancer	Medications
Colon	Sedatives
Hepatobiliary	Antibiotics
Hematologic	Nonsteroidal anti-inflammatory drugs
Lung	Serotonin reuptake inhibitors
Breast	Metformin
Genitourinary	Levodopa
Ovarian	Angiotensin-converting enzyme inhibitors
Prostate	Other drugs
Gastrointestinal disorders	Disorders of the mouth and teeth
Difficulty swallowing	Caries
Malabsorption	Dysgeusia
Peptic ulcer	
Inflammatory bowel disease	
Pancreatitis	Age-related factors
Obstruction/constipation	Physiologic changes
Pernicious anemia	Visual impairment
Endocrine and metabolic	Decreased taste and smell
Hyperthyroidism	Functional disabilities
Diabetes mellitus	
Pheochromocytoma	Neurologic
Adrenal insufficiency	Stroke
Cardiac disorders	Parkinson's disease
Chronic ischemia	Neuromuscular disorders
Chronic congestive heart failure	Dementia
Respiratory disorders	
Emphysema	Social
Chronic obstructive pulmonary disease	Isolation
Renal insufficiency	Poverty
Rheumatologic disease	
Infections	Psychiatric and behavioral
HIV	Depression
Tuberculosis	Anxiety
Parasitic infection	Paranoia
Subacute bacterial endocarditis	Bereavement
	Alcoholism
	Eating disorders
	Increased activity or exercise
	Idiopathic

and pulmonary diseases cause UWL through increased metabolic demand and decreased appetite and caloric intake. Repeated surgeries may lead to weight loss because of reduced caloric intake and increased metabolic demands resulting from a systemic inflammatory response. Uremia produces nausea, anorexia, and vomiting. Connective tissue diseases may increase metabolic demand and disrupt nutritional balance. As the incidence of diabetes mellitus increases with aging, the associated glucosuria can contribute to weight loss. Hyperthyroidism in the elderly may have less prominent sympathomimetic features and may present as “apathetic hyperthyroidism” or T₃ toxicosis (Chap. 382).

Neurologic injuries such as stroke, quadriplegia, and multiple sclerosis may lead to visceral and autonomic dysfunction that can impair caloric intake. Dysphagia from these neurologic insults is a common mechanism. Functional disability that compromises activities of daily living (ADLs) is a common cause of undernutrition in the elderly. Visual impairment from ophthalmic or central nervous system disorders such as a tremor can limit the ability of people to prepare and eat meals. UWL may be one of the earliest manifestations of Alzheimer’s dementia.

Isolation and depression are significant causes of UWL that may manifest as an inability to care for oneself, including nutritional needs. A cytokine-mediated inflammatory metabolic cascade can be both a cause of and a manifestation of depression. Bereavement can be a cause of UWL and, when present, is often more pronounced in men. More intense forms of mental illness such as paranoid disorders may

lead to delusions about food and cause weight loss. Alcoholism can be a significant source of weight loss and malnutrition.

Elderly persons living in poverty may have to choose whether to purchase food or use the money for other expenses, including medications. Screening questions can probe whether patients have run out of food or whether they routinely purchase less than they need. Institutionalization is an independent risk factor, as up to 30–50% of nursing home patients have inadequate food intake.

Medications can cause anorexia, nausea, vomiting, gastrointestinal distress, diarrhea, dry mouth, and changes in taste. This is particularly an issue in the elderly, many of whom take five or more medications.

ASSESSMENT

The four major manifestations of UWL are (1) anorexia (loss of appetite), (2) sarcopenia (loss of muscle mass), (3) cachexia (a syndrome that combines weight loss, loss of muscle and adipose tissue, anorexia, and weakness), and (4) dehydration. The current obesity epidemic adds complexity, as excess adipose tissue can mask the development of sarcopenia and delay awareness of the development of cachexia. If it is not possible to measure weight directly, a change in clothing size, corroboration of weight loss by a relative or friend, and a numeric estimate of weight loss provided by the patient are suggestive of true weight loss.

Initial assessment includes a comprehensive history and physical, a complete blood count, tests of liver enzyme levels, C-reactive protein, erythrocyte sedimentation rate, renal function studies, thyroid function tests, chest radiography, and an abdominal ultrasound (Table 47-2). Age-, sex-, and risk factor-specific cancer screening tests, such as mammography and colonoscopy, should be performed (Chap. 70). Patients at risk should have HIV testing. All elderly patients with weight loss should undergo screening for dementia and depression by using instruments such as the Mini-Mental State Examination and the Geriatric Depression Scale, respectively (Chap. 477). The Mini Nutritional Assessment (www.mna-elderly.com) and the Nutrition Screening Initiative (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1694757/>) are also available for the nutritional assessment of elderly patients. Almost all patients with a malignancy and >90% of those with other organic diseases have at least one laboratory abnormality. In patients presenting with substantial UWL, major organic and malignant diseases are unlikely when

TABLE 47-2 Assessment and Testing for Involuntary Weight Loss

Indications	Laboratory
5% weight loss in 30 d	Complete blood count
10% weight loss in 180 d	Comprehensive electrolyte and metabolic panel, including liver and renal function tests
Body mass index <21	Thyroid function tests
25% of food left uneaten after 7 d	Erythrocyte sedimentation rate
Change in fit of clothing	C-reactive protein
Change in appetite, smell, or taste	Ferritin
Abdominal pain, nausea, vomiting, diarrhea, constipation, dysphagia	HIV testing, if indicated
Assessment	Radiology
Complete physical examination, including dental evaluation	Chest x-ray Abdominal ultrasound
Medication review	
Recommended cancer screening	
Mini-Mental State Examination ^a	
Mini-Nutritional Assessment ^a	
Nutrition Screening Initiative ^a	
Simplified Nutritional Assessment Questionnaire ^a	
Observation of eating ^a	
Activities of daily living ^a	
Instrumental activities of daily living ^a	

^aMay be more specific to assess weight loss in the elderly.

a baseline evaluation is completely normal. Careful follow-up rather than undirected testing is advised because the prognosis of weight loss of undetermined cause is generally favorable.

TREATMENT

Unintentional Weight Loss

The first priority in managing weight loss is to identify and treat the underlying causes. Treatment of underlying metabolic, psychiatric, infectious, or other systemic disorders may be sufficient to restore weight and functional status gradually. Medications that cause nausea or anorexia should be withdrawn or changed, if possible. For those with unexplained UWL, oral nutritional supplements such as high-energy drinks sometimes reverse weight loss. Advising patients to consume supplements between meals rather than with a meal may help minimize appetite suppression and facilitate increased overall intake. Orexigenic, anabolic, and anticytokine agents are under investigation. In selected patients, the antidepressant mirtazapine results in a significant increase in body weight, body fat mass, and leptin concentration. Patients with wasting conditions who can comply with an appropriate exercise program gain muscle protein mass, strength, and endurance and may be more capable of performing ADLs.

ACKNOWLEDGMENT

The author is grateful to Russell G. Robertson, MD, for contributions to this chapter in prior editions.

FURTHER READING

- ALIBHAI SM et al: An approach to the management of unintentional weight loss in elderly people. *CMAJ* 172:773, 2005.
- GADDEY HL, HOLDER K: Unintentional weight loss in older adults. *Am Fam Physician* 89:718, 2014.
- MCMINN J et al: Investigation and management of unintentional weight loss in older adults. *BMJ* 342:d1732, 2011.
- NICHOLSON BD et al: Prioritising primary care patients with unexpected weight loss for cancer investigation. *BMJ* 370:m2651, 2020.
- VANDERSCHUEREN S et al: The diagnostic spectrum of unintentional weight loss. *Eur J Intern Med* 16:160, 2005.
- WONG CJ: Involuntary weight loss. *Med Clin North Am* 98:625, 2014.

SOURCES OF GASTROINTESTINAL BLEEDING

Upper Gastrointestinal Sources of Bleeding • PEPTIC ULCERS Peptic ulcers are the most common cause of upper GIB (UGIB), accounting for ~50% of UGIB hospitalizations. Features of an ulcer at endoscopy provide important prognostic information that guides subsequent management decisions (Fig. 48-1). Approximately 20% of patients with bleeding ulcers have the highest-risk findings of active bleeding or a nonbleeding visible vessel; one-third of such patients have further bleeding that requires urgent surgery if they are treated conservatively. These patients benefit from endoscopic therapy such as bipolar electrocoagulation, heater probe, injection therapy (e.g., absolute alcohol, 1:10,000 epinephrine), and/or clips with reductions in bleeding, hospital stay, mortality, and costs. In contrast, patients with clean-based ulcers have rates of serious recurrent bleeding approaching zero. If stable with no other reason for hospitalization, such patients may be discharged home after endoscopy.

Randomized controlled trials document that high-dose, constant-infusion IV proton pump inhibitor (PPI) (80-mg bolus and 8-mg/h infusion), designed to sustain intragastric pH >6 and enhance clot stability, decreases further bleeding and mortality in patients with high-risk ulcers (active bleeding, nonbleeding visible vessel, adherent clot) when given after endoscopic therapy. Meta-analysis of randomized trials indicates that high-dose intermittent PPIs are noninferior to constant-infusion PPI therapy and thus may be substituted. Patients with lower-risk findings (flat pigmented spot or clean base) do not require endoscopic therapy and receive standard doses of oral PPI.

Approximately 10–50% of patients with bleeding ulcers rebleed within the next year if no preventive strategies are employed. Prevention of recurrent bleeding focuses on the three main factors in ulcer pathogenesis, *Helicobacter pylori*, nonsteroidal anti-inflammatory drugs (NSAIDs), and acid. Eradication of *H. pylori* in patients with bleeding ulcers decreases rebleeding rates to <5%. If a bleeding ulcer develops in a patient taking NSAIDs, the NSAIDs should be discontinued. If NSAIDs must be given, a cyclooxygenase (COX)-2 selective NSAID plus a PPI is recommended, based on results of a randomized trial. Patients with established cardiovascular disease who develop bleeding ulcers while taking low-dose aspirin for secondary prevention should restart aspirin as soon as possible after their bleeding episode (1–7 days). A randomized trial showed that immediate reinstitution of aspirin was associated with a lower 8-week mortality compared to not restarting aspirin (1% vs 13%; hazard ratio, 0.2; 95% CI, 0.1–0.6). In contrast, aspirin probably should be discontinued in most patients taking aspirin for primary prevention of cardiovascular events who develop UGIB. Patients with bleeding ulcers unrelated to *H. pylori* or NSAIDs should remain on PPI therapy indefinitely given a 42% incidence of rebleeding at 7 years without protective therapy. **Peptic ulcers are discussed in Chap. 324.**

MALLORY-WEISS TEARS Mallory-Weiss tears account for ~2–10% of UGIB hospitalizations. The classic history is vomiting, retching, or coughing preceding hematemesis, especially in an alcoholic patient. Bleeding from these tears, which are usually on the gastric side of the gastroesophageal junction, stops spontaneously in ~80–90% of patients and recurs in only 0–10%. Endoscopic therapy is indicated for actively bleeding Mallory-Weiss tears. **Mallory-Weiss tears are discussed in Chap. 323.**

ESOPHAGEAL VARICES The proportion of UGIB hospitalizations due to varices varies widely, from ~2–40%, depending on the population. Patients with variceal hemorrhage have poorer outcomes than patients with other sources of UGIB. Esophageal varices are treated with endoscopic ligation and an IV vasoactive medication (octreotide, somatostatin, vaptoreotide, terlipressin) for 2–5 days. Combination of endoscopic and medical therapy is superior to either therapy alone in decreasing rebleeding. Over the long term, treatment with nonselective beta blockers plus endoscopic ligation is recommended because the combination is more effective than either alone in reduction of recurrent esophageal variceal bleeding. Transjugular intrahepatic portosystemic shunt (TIPS) is recommended in patients who have persistent or

48

Gastrointestinal Bleeding

Loren Laine

Gastrointestinal bleeding (GIB) presents as either overt or occult bleeding. *Overt GIB* is manifested by *hematemesis*, vomitus of red blood or “coffee-grounds” material; *melena*, black, tarry stool; and/or *hematochezia*, passage of red or maroon blood from the rectum. In the absence of overt bleeding, *occult GIB* may present with *symptoms of blood loss or anemia* such as lightheadedness, syncope, angina, or dyspnea; with iron-deficiency anemia; or a positive fecal occult blood test on colorectal cancer screening. GIB is also categorized by the site of bleeding as upper, from the esophagus, stomach, or duodenum; lower, from the colon; small intestinal; or obscure GIB if the source is unclear.

GIB is the most common gastrointestinal condition leading to hospitalization in the United States, accounting for ~513,000 admissions and \$5 billion in direct costs annually. The case fatality of patients hospitalized with GIB is ~2% in the United States. Patients generally die from decompensation of other underlying illnesses rather than exsanguination.

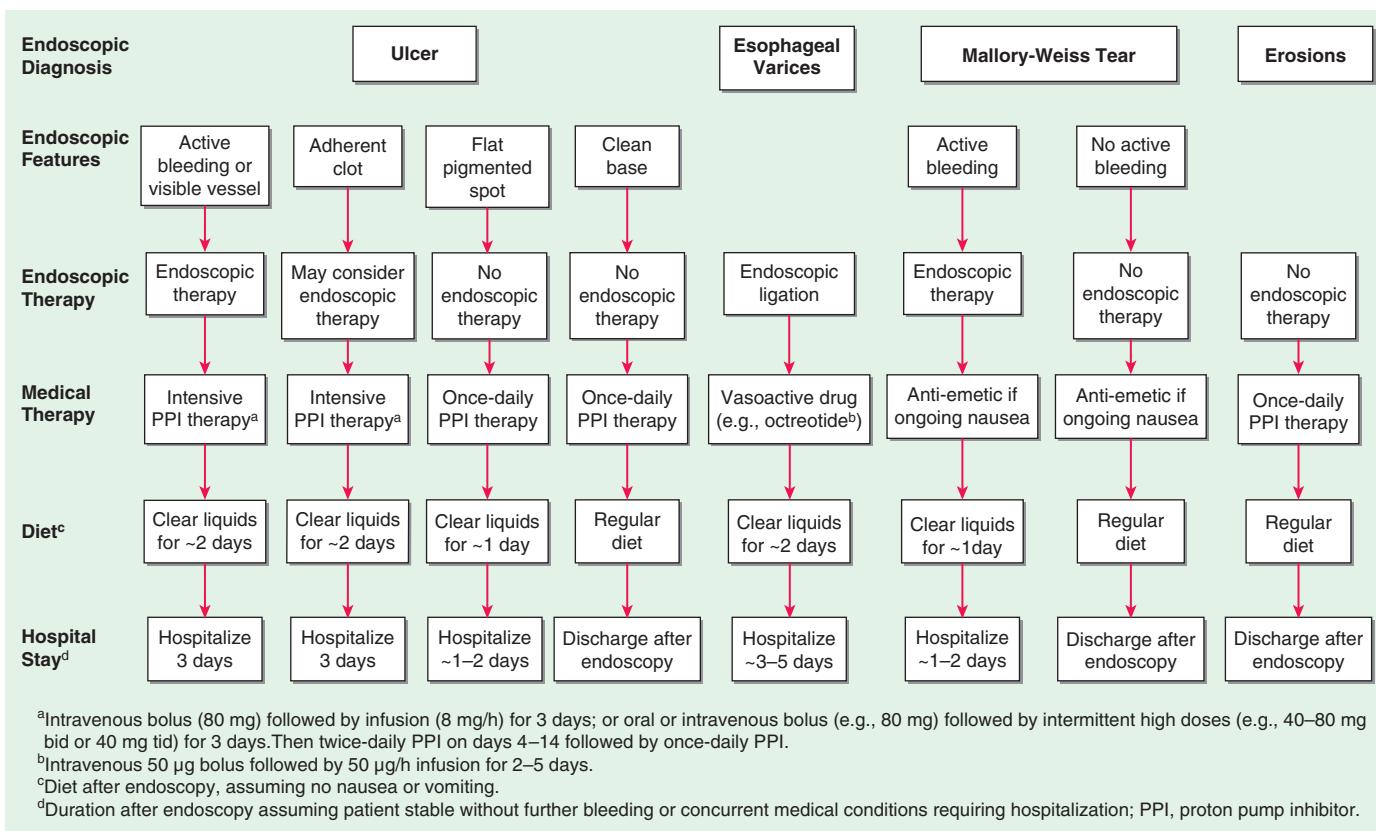


FIGURE 48-1 Suggested algorithm for patients with acute upper gastrointestinal bleeding based on endoscopic findings.

recurrent bleeding despite endoscopic and medical therapy. TIPS also should be considered in the first 1–2 days of hospitalization for acute variceal bleeding in patients with advanced liver disease (Child-Pugh class B, Child-Pugh class C with score 10–13), because randomized trials show significant decreases in rebleeding and mortality compared with standard endoscopic and medical therapy.

Portal hypertension is also responsible for bleeding from gastric varices, varices in the small and large intestine, and portal hypertensive gastropathy and enterocolopathy. Bleeding gastric varices are treated with endoscopic injection of tissue adhesive (e.g., *n*-butyl cyanoacrylate), if available; if not, TIPS is performed.

EROSIVE DISEASE Erosions are endoscopically visualized breaks that are confined to the mucosa and do not cause major bleeding because arteries and veins are not present in the mucosa. Erosions in the esophagus, stomach, or duodenum commonly cause mild UGIB, with erosive gastritis and duodenitis accounting for perhaps ~10–15% and erosive esophagitis (primarily due to gastroesophageal reflux disease) accounting for ~1–10% of UGIB hospitalizations. The most important cause of gastric and duodenal erosions is NSAID use: ~50% of patients who chronically ingest NSAIDs may have gastric erosions. Other potential causes of gastric erosions include alcohol intake, *H. pylori* infection, and stress-related mucosal injury.

Stress-related gastric mucosal injury occurs only in extremely sick patients, such as those with serious trauma, major surgery, burns covering more than one-third of the body surface area, major intracranial disease, or severe medical illness (e.g., ventilator dependence, coagulopathy). Severe bleeding should not develop unless ulceration occurs. The mortality rate in these patients is high because of their serious underlying illnesses.

The incidence of bleeding from stress-related gastric mucosal injury has decreased dramatically in recent years, most likely due to better care of critically ill patients. A recent double-blind placebo-controlled randomized trial in 3282 intensive care patients with risk factors for GIB showed a small benefit of PPI in clinically important bleeding (2.5% vs 4.2%) without a difference in mortality or infections (e.g.,

Clostridium difficile, pneumonia). Thus, pharmacologic prophylaxis for bleeding has limited benefit but may be considered in the high-risk patients mentioned above. Meta-analyses of randomized trials suggest PPIs are more effective than H₂-receptor antagonists in reduction of overt and clinically important UGIB without differences in mortality or nosocomial pneumonia.

OTHER CAUSES Less common causes of UGIB include neoplasms, vascular ectasias (including hereditary hemorrhagic telangiectasias [Osler-Weber-Rendu] and gastric antral vascular ectasia ["watermelon stomach"]), Dieulafoy's lesion (in which an aberrant vessel in the mucosa bleeds from a pinpoint mucosal defect), prolapse gastropathy (prolapse of proximal stomach into esophagus with retching, especially in alcoholics), aortoenteric fistulas, and hemobilia or hemosuccus pancreaticus (bleeding from the bile duct or pancreatic duct).

Small-Intestinal Sources of Bleeding Patients without a source of GIB identified on upper endoscopy and colonoscopy were previously labeled as having obscure GIB. With the advent of improved diagnostic modalities, ~75% of GIB previously labeled obscure is now estimated to originate in the small intestine beyond the extent of a standard upper endoscopic exam. Small-intestinal GIB may account for ~5% of GIB cases. The most common causes in adults include vascular ectasias, neoplasm (e.g., gastrointestinal stromal tumor, carcinoid, adenocarcinoma, lymphoma, metastases), and NSAID-induced erosions and ulcers. Meckel's diverticulum is the most common cause of significant small-intestinal GIB in children, decreasing in frequency as a cause of bleeding with age. Other less common causes of small-intestinal GIB include Crohn's disease, infection, ischemia, vasculitis, small-bowel varices, diverticula, intussusception, Dieulafoy's lesions, aortoenteric fistulas, and duplication cysts.

Small-intestinal vascular ectasias are treated with endoscopic therapy, if possible, based on observational studies suggesting initial efficacy. However, rebleeding is common: 45% over a mean follow-up of 26 months in a systematic review. Estrogen/progesterone compounds are not recommended because a multicenter double-blind trial found no benefit in prevention of recurrent bleeding. Octreotide is used,

based on positive results from case series but no randomized trials. A randomized trial reported significant benefit of thalidomide and awaits further confirmation. Other isolated lesions, such as tumors, generally require surgical resection.

Colonic Sources of Bleeding Hemorrhoids are probably the most common cause of lower GIB (LGIB); anal fissures also cause minor bleeding and pain. If these local anal processes, which rarely require hospitalization, are excluded, the most common cause of LGIB in adults is diverticulosis. Other causes include vascular ectasias (especially in the proximal colon of patients >70 years), neoplasms (primarily adenocarcinoma), colitis (ischemic, infectious, Crohn's or ulcerative colitis, NSAID-induced colitis or ulcers), postpolypectomy bleeding, and radiation proctopathy. Rarer causes include solitary rectal ulcer syndrome, varices (most commonly rectal), lymphoid nodular hyperplasia, vasculitis, trauma, and aortocolic fistulas. In children and adolescents, the most common colonic causes of significant GIB are inflammatory bowel disease and juvenile polyps.

Diverticular bleeding is abrupt in onset, usually painless, sometimes massive, and often from the right colon; chronic or occult bleeding is not characteristic. Case series from the United States and Europe suggest colonic diverticula stop bleeding spontaneously in ≥90% of patients, with rebleeding on long-term follow-up as low as ~15% over 4–5 years. Rebleeding is substantially higher in reports from Asia. Case series suggest endoscopic therapy may decrease recurrent bleeding in the uncommon case when colonoscopy identifies the specific bleeding diverticulum. When diverticular bleeding is found at angiography, transcatheter arterial embolization by superselective technique stops bleeding in a majority of patients. Segmental surgical resection is recommended for persistent or refractory diverticular bleeding.

Bleeding from colonic vascular ectasias may be overt or occult; it tends to be chronic and only occasionally hemodynamically significant. Endoscopic hemostatic therapy may be used in the treatment of vascular ectasias, as well as discrete bleeding ulcers and post-polypectomy bleeding. Transcatheter arterial embolization also may be attempted for persistent bleeding from vascular ectasias and other discrete lesions. Surgical therapy is generally required for major persistent or recurrent bleeding from colonic sources that cannot be treated medically, endoscopically, or angiographically. Patients with Heyde's syndrome (bleeding vascular ectasias and aortic stenosis) appear to benefit from aortic valve replacement.

APPROACH TO THE PATIENT

Gastrointestinal Bleeding

INITIAL ASSESSMENT

Measurement of the heart rate and blood pressure is the best way to initially assess a patient with GIB. Clinically significant bleeding leads to postural changes in heart rate or blood pressure, tachycardia, and, finally, recumbent hypotension. In contrast, hemoglobin does not fall immediately with acute GIB, due to proportionate reductions in plasma and red cell volumes ("people bleed whole blood"). Thus, hemoglobin may be normal or only minimally decreased at initial presentation of a severe bleeding episode. As extravascular fluid enters the vascular space to restore volume, the hemoglobin falls, but this process may take up to 72 h. Transfusion is recommended when the hemoglobin drops below 7 g/dL, based on a large randomized trial showing this restrictive transfusion strategy decreases rebleeding and death in acute UGIB compared with a transfusion threshold of 9 g/dL. Patients with slow, chronic GIB may have very low hemoglobin values despite normal blood pressure and heart rate. With the development of iron-deficiency anemia, the mean corpuscular volume is low and red blood cell distribution width is increased.

DIFFERENTIATION OF UGIB FROM LGIB

Hematemesis indicates an UGIB source. Melena indicates blood has been present in the gastrointestinal (GI) tract for ≥14 h and as

long as 3–5 days. The more proximal the bleeding site, the more likely melena will occur. Hematochezia usually represents a lower GI source of bleeding, although an upper GI lesion may bleed so briskly that blood transits the bowel before melena develops. When hematochezia is the presenting symptom of UGIB, it is associated with hemodynamic instability and dropping hemoglobin. Bleeding lesions of the small bowel may present as melena or hematochezia. Other clues to UGIB include hyperactive bowel sounds and an elevated blood urea nitrogen (due to volume depletion and blood proteins absorbed in the small intestine).

A nonbloody nasogastric aspirate may be seen in ~15% of patients with UGIB who present with clinically serious hematochezia. A bile-stained appearance does not exclude UGIB because reports of bile in the aspirate are incorrect in ~50% of cases. Testing of aspirates that are not grossly bloody for occult blood is not useful.

EVALUATION AND MANAGEMENT OF UGIB (FIG. 48-1)

Initial Risk Assessment Baseline characteristics predictive of rebleeding and death include hemodynamic compromise (tachycardia or hypotension), increasing age, and comorbidities. Risk assessment tools may be used to identify patients with very low risk. Discharge from the emergency room with outpatient management has been suggested for patients with a Glasgow-Blatchford score (possible range 0–23, Table 48-1) of 0–1 because only ~1% of patients who require transfusion, require hemostatic intervention, or die have a score of 0–1.

Pre-Endoscopic Medications PPI infusion may be considered at presentation; it decreases high-risk ulcer stigmata (e.g., active bleeding) and need for endoscopic therapy but does not improve clinical outcomes such as further bleeding, surgery, or death. The promotility agent erythromycin, 250 mg intravenously ~30–90 min before endoscopy, is suggested to improve visualization at endoscopy, thereby reducing the need for repeat endoscopy and hospital stay. Cirrhotic patients presenting with UGIB should be given an antibiotic (e.g., ceftriaxone) and IV vasoactive medication (e.g., octreotide) upon presentation. Antibiotics decrease bacterial infections, rebleeding, and mortality, and vasoactive medications may improve control of bleeding in the 12 h after presentation.

Endoscopy Upper endoscopy should be performed within 24 h in most patients hospitalized with UGIB whether they have clinical features predicting low risk or high risk of further bleeding

TABLE 48-1 Glasgow-Blatchford Score

RISK FACTORS AT ADMISSION	SCORE
Blood urea nitrogen (mg/dL)	
18.2 to <22.4	2
22.4 to <28.0	3
28.0 to <70.0	4
≥70.0	6
Hemoglobin (g/dL)	
12.0 to <13.0 (men); 10.0 to <12.0 (women)	1
10.0 to <12.0 (men)	3
<10.0	6
Systolic blood pressure (mmHg)	
100–109	1
90–99	2
<90	3
Heart rate (beats per minute)	
≥100	1
Melena	1
Syncope	2
Hepatic disease	2
Cardiac failure	2

and death. Even in high-risk patients, more urgent endoscopy (performed within 6 h of gastroenterology consultation) does not improve clinical outcomes. Early endoscopy in low-risk patients (e.g., hemodynamically stable without severe comorbidities) identifies low-risk findings (e.g., clean-based ulcers, erosions, nonbleeding Mallory-Weiss tears) that allow discharge in ≥40% of patients, thereby reducing hospital stay and costs. Patients with high-risk endoscopic findings (e.g., varices, ulcers with active bleeding or a visible vessel) benefit from hemostatic therapy at endoscopy.

EVALUATION AND MANAGEMENT OF LGIB (FIG. 48-2)

Patients with hematochezia and hemodynamic instability should have upper endoscopy to rule out an upper GI source before evaluation of the lower GI tract.

Colonoscopy after an oral lavage solution is the procedure of choice in most patients admitted with LGIB unless bleeding is too massive, in which case angiography is recommended. Computed tomography (CT) angiography is often suggested prior to angiography to document evidence and location of active bleeding. Sigmoidoscopy is used primarily in patients <40 years old with minor bleeding. In patients with no source identified on colonoscopy, imaging studies may be employed. ^{99m}Tc -labeled red cell scan allows repeated imaging for up to 24 h and may identify the general location of bleeding. However, CT angiography is increasingly used instead because it is likely superior and more readily available. In active LGIB, angiography can detect the site of bleeding (extravasation of contrast into the gut) and permits treatment with transcatheter arterial embolization.

EVALUATION AND MANAGEMENT OF SMALL-INTESTINAL OR OBSCURE GIB

In patients with massive bleeding suspected to be from the small intestine, current guidelines suggest angiography as the initial test, with CT angiography or ^{99m}Tc -labeled red cell scan prior to

angiography if the patient's clinical status permits. For others, repeat upper and lower endoscopy may be considered as the initial evaluation because second-look procedures identify a source in up to ~25% of upper endoscopies and colonoscopies; a push enteroscopy, usually performed with a pediatric colonoscope to inspect the entire duodenum and proximal jejunum, may be substituted for a repeat standard upper endoscopy. If second-look procedures are negative, evaluation of the entire small intestine is performed, usually with video capsule endoscopy. A systematic review of comparative studies showed the yield of "clinically significant findings" to be greater with capsule than push enteroscopy (56% vs 26%) or small bowel barium radiography (42% vs 6%). However, capsule endoscopy does not allow full visualization of the small intestine, tissue sampling, or application of therapy.

CT enterography may be used initially instead of video capsule in patients with possible small bowel narrowing (e.g., stricture, prior surgery or radiation, Crohn's disease) and may follow a negative video capsule for suspected small-intestinal GIB, given its higher sensitivity for small-intestinal masses.

If capsule endoscopy is positive, management is dictated by the finding. If capsule endoscopy is negative, clinically stable patients may be observed and treated with iron if iron deficiency is present, while those with ongoing bleeding (e.g., need for transfusions) undergo further testing. A second capsule endoscopy may be considered because it is reported to identify a source in up to ~50% of cases. "Deep" enteroscopy (double-balloon, single-balloon, or spiral enteroscopy) is commonly the next test after capsule endoscopy for clinically important GIB documented or suspected to be from the small intestine because it allows the endoscopist to examine, obtain specimens from, and provide therapy to much or all of the small intestine. Other imaging techniques sometimes used in evaluation of obscure GIB include ^{99m}Tc -labeled red blood cell scintigraphy, CT angiography, angiography, and ^{99m}Tc -pertechnetate scintigraphy for Meckel's

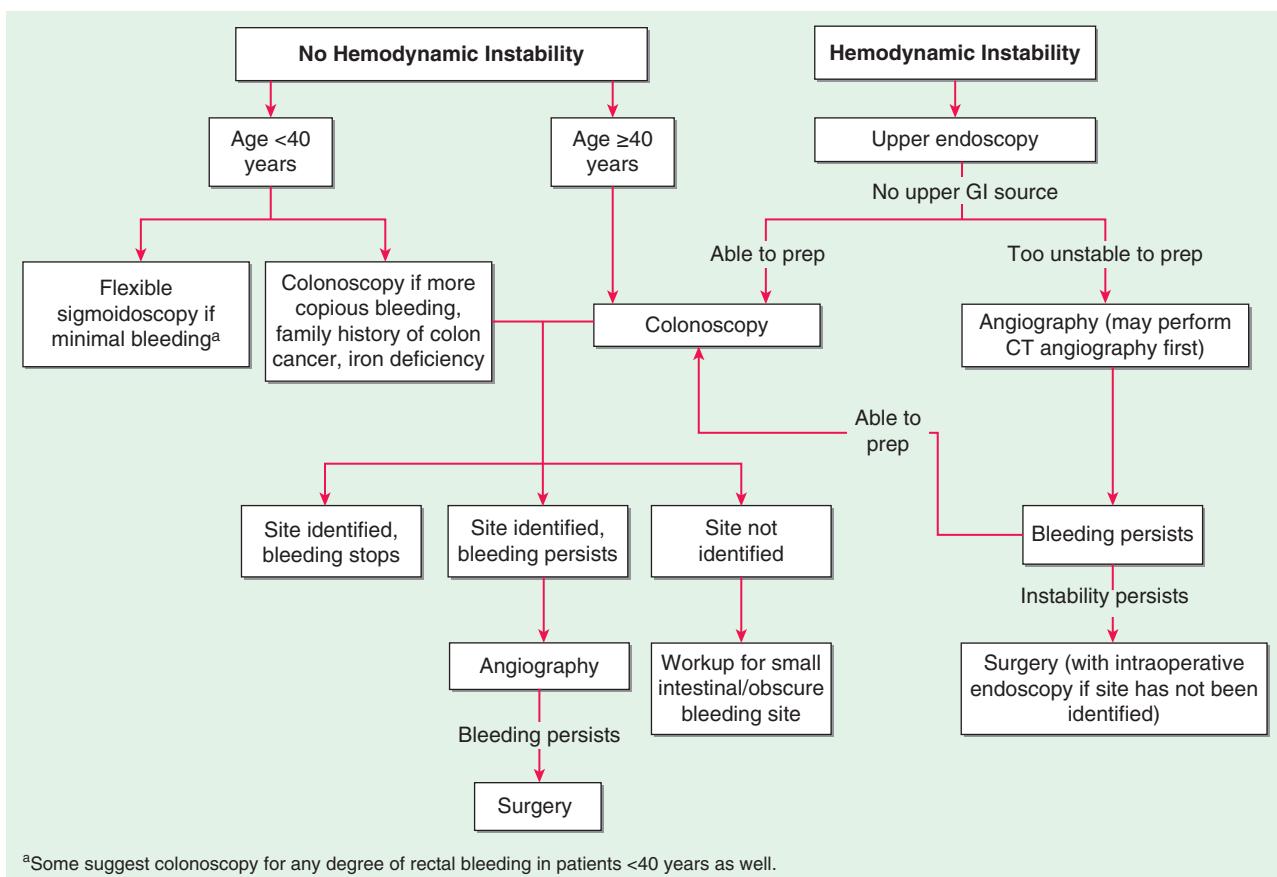


FIGURE 48-2 Suggested algorithm for patients with acute lower gastrointestinal bleeding.

diverticulum (especially in young patients). If all tests are unrevealing, intraoperative endoscopy is indicated in patients with severe recurrent or persistent bleeding requiring repeated transfusions.

POSITIVE FECAL OCCULT BLOOD TEST

Fecal occult blood testing is recommended only for colorectal cancer screening, beginning at age 45–50 years in average-risk adults. A positive test necessitates colonoscopy. If evaluation of the colon is negative, further workup is not recommended unless iron-deficiency anemia or GI symptoms are present.

FURTHER READING

- GARCIA-TSAO G et al: Portal hypertensive bleeding in cirrhosis: Risk stratification, diagnosis, and management: 2016 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology* 65:310, 2017.
- GURUDU SR et al: The role of endoscopy in the management of suspected small-bowel bleeding. *Gastrointest Endosc* 85:22, 2017.
- KRAG M et al: Pantoprazole in patients at risk for gastrointestinal bleeding in the ICU. *N Engl J Med* 379:2199, 2018.
- LAINE L et al: ACG clinical guideline: Upper gastrointestinal and ulcer bleeding. *Am J Gastroenterol* 116:899, 2021.
- LAU JYW et al: Timing of endoscopy for acute upper gastrointestinal bleeding. *N Engl J Med* 382:1299, 2020.
- PEERY AF et al: Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: Update 2018. *Gastroenterology* 156:254, 2019.
- STANLEY AJ, LAINE L: Management of acute upper gastrointestinal bleeding. *BMJ* 364:i536, 2019.
- STRATE LL, GRALNEK KM: ACG clinical guideline: Management of patients with acute lower gastrointestinal bleeding. *Am J Gastroenterol* 111:459, 2016.
- VILLANEUVA C et al: Transfusion strategies for acute upper gastrointestinal bleeding. *N Engl J Med* 368:11, 2013.

49

Jaundice

Savio John, Daniel S. Pratt



Jaundice is a yellowish discoloration of body tissues resulting from the deposition of bilirubin. Tissue deposition of bilirubin occurs only in the presence of serum hyperbilirubinemia and is a sign of either liver disease or, less often, a hemolytic disorder or disorder of bilirubin metabolism. The degree of serum bilirubin elevation can be estimated by physical examination. Slight increases in serum bilirubin level are best detected by examining the sclerae for icterus. Sclerae have a particular affinity for bilirubin due to their high elastin content, and the presence of scleral icterus indicates a serum bilirubin level of at least 51 µmol/L (3 mg/dL). The ability to detect scleral icterus is made more difficult if the examining room has fluorescent lighting. If the examiner suspects scleral icterus, a second site to examine is underneath the tongue. As serum bilirubin levels rise, the skin will eventually become yellow in light-skinned patients and even green if the process is long-standing; the green color is produced by oxidation of biliverdin.

The differential diagnosis for yellowing of the skin is limited. In addition to jaundice, it includes carotenoderma; the use of drugs including quinacrine, sunitinib, and sorafenib; and excessive exposure to phenols. Carotenoderma, a yellow coloring of the skin, is associated with diabetes, hypothyroidism, and anorexia nervosa, but most commonly, it is caused by the ingestion of an excessive amounts of vegetables and fruits such as carrots, leafy vegetables, squash, peaches, and

oranges that contain carotene. In jaundice, the yellow coloration of the skin is uniformly distributed over the body, whereas in carotenoderma, the pigment is concentrated on the palms, soles, forehead, and nasolabial folds. Carotenoderma can be distinguished from jaundice by the sparing of the sclerae. Quinacrine causes a yellow discoloration of the skin in 4–37% of patients treated with it. It has also been reported with the use of the tyrosine kinase inhibitors sunitinib and sorafenib.

Another sensitive indicator of increased serum bilirubin is darkening of the urine, which is due to the renal excretion of conjugated bilirubin. Patients often describe their urine as tea- or cola-colored. Bilirubinuria indicates an elevation of the direct serum bilirubin fraction and, therefore, the presence of liver or biliary disease.

Serum bilirubin levels increase when an imbalance exists between bilirubin production and clearance. A logical evaluation of the patient who is jaundiced requires an understanding of bilirubin production and metabolism.

PRODUCTION AND METABOLISM OF BILIRUBIN

(See Chap. 338) Bilirubin, a tetrapyrrole pigment, is a breakdown product of heme (ferroprotoporphyrin IX). About 80–85% of the 4 mg/kg body weight of bilirubin produced each day is derived from the breakdown of hemoglobin in senescent red blood cells. The remainder comes from prematurely destroyed erythroid cells in bone marrow and from the turnover of hemoproteins such as myoglobin and cytochromes found in tissues throughout the body.

The formation of bilirubin occurs in reticuloendothelial cells, primarily in the spleen and liver. The first reaction, catalyzed by the microsomal enzyme heme oxygenase, oxidatively cleaves the α bridge of the porphyrin group and opens the heme ring. The end products of this reaction are biliverdin, carbon monoxide, and iron. The second reaction, catalyzed by the cytosolic enzyme biliverdin reductase, reduces the central methylene bridge of biliverdin and converts it to bilirubin. Bilirubin formed in the reticuloendothelial cells is virtually insoluble in water due to tight internal hydrogen bonding between the water-soluble moieties of bilirubin—that is, the bonding of the propionic acid carboxyl groups of one dipyrrrolid half of the molecule with the imino and lactam groups of the opposite half. This configuration blocks solvent access to the polar residues of bilirubin and places the hydrophobic residues on the outside. To be transported in blood, bilirubin must be solubilized. Solubilization is accomplished by the reversible, noncovalent binding of bilirubin to albumin. Unconjugated bilirubin bound to albumin is transported to the liver. There, the bilirubin—but not the albumin—is taken up by hepatocytes via a process that at least partly involves carrier-mediated membrane transport. No specific bilirubin transporter has yet been identified (Chap. 338, Fig. 338-1).

After entering the hepatocyte, unconjugated bilirubin is bound in the cytosol to several proteins including proteins in the glutathione-S-transferase superfamily. These proteins serve both to reduce efflux of bilirubin back into the serum and to present the bilirubin for conjugation. In the endoplasmic reticulum, bilirubin is made aqueous soluble by conjugation to glucuronic acid, a process that disrupts the hydrophobic internal hydrogen bonds and yields bilirubin monoglucuronide and diglucuronide. The conjugation of glucuronic acid to bilirubin is catalyzed by bilirubin uridine diphosphate-glucuronosyl transferase (UDPGT). The now-hydrophilic bilirubin conjugates diffuse from the endoplasmic reticulum to the canalicular membrane, where bilirubin monoglucuronide and diglucuronide are actively transported into canalicular bile by an energy-dependent mechanism involving the multidrug resistance-associated protein 2 (MRP2). A portion of bilirubin glucuronides is transported into the sinusoids and portal circulation by MRP3 and is subjected to reuptake into the hepatocyte by the sinusoidal organic anion transport protein 1B1 (OATP1B1) and OATP1B3. The conjugated bilirubin excreted into bile drains into the duodenum and passes unchanged through the proximal small bowel. Conjugated bilirubin is not reabsorbed by the intestinal mucosa due to its hydrophilicity and increased molecular size. When the conjugated bilirubin reaches the distal ileum and colon, it is hydrolyzed to unconjugated bilirubin by bacterial β-glucuronidases.

The unconjugated bilirubin is reduced by normal gut bacteria to form a group of colorless tetrapyrroles called *urobilinogens* and other products, the nature and relative amounts of which depend on the bacterial flora. About 80–90% of these products are excreted in feces, either unchanged or oxidized to orange derivatives called *urobilins*. The remaining 10–20% of the urobilinogens undergo enterohepatic cycling. A small fraction (usually <3 mg/dL) escapes hepatic uptake, filters across the renal glomerulus, and is excreted in urine. Increased urinary excretion of urobilinogen can be due to increased bilirubin production, increased hepatic reabsorption of urobilinogen from the colon, or decreased hepatic clearance of urobilinogen.

MEASUREMENT OF SERUM BILIRUBIN

The terms *direct* and *indirect* bilirubin—that is, conjugated and unconjugated bilirubin, respectively—are based on the original van den Bergh reaction. This assay, or a variation of it, is still used in most clinical chemistry laboratories to determine the serum bilirubin level. In this assay, bilirubin is exposed to diazotized sulfanilic acid and splits into two relatively stable dipyrrylmethene azopigments that absorb maximally at 540 nm, allowing photometric analysis. The direct fraction is that which reacts with diazotized sulfanilic acid in the absence of an accelerator substance such as alcohol. The direct fraction provides an approximation of the conjugated bilirubin level in serum. The *total* serum bilirubin is the amount that reacts after the addition of alcohol. The indirect fraction is the difference between the total and the direct bilirubin levels and provides an estimate of the unconjugated bilirubin in serum. Unconjugated bilirubin also reacts with diazo reagents, albeit slowly, even when the accelerator is absent. Thus, the calculated indirect bilirubin may underestimate the true amount of unconjugated bilirubin in circulation.

With the van den Bergh method, the normal serum bilirubin concentration usually is between 17 and 26 µmol/L (1 and 1.5 mg/dL). Total serum bilirubin concentrations are between 3.4 and 15.4 µmol/L (0.2 and 0.9 mg/dL) in 95% of a normal population. Unconjugated hyperbilirubinemia is present when the direct fraction is <15% of the total serum bilirubin. The presence of even limited amounts of true conjugated bilirubin in serum suggests significant hepatobiliary pathology. As conjugated hyperbilirubinemia is always associated with bilirubinuria (except in the presence of delta bilirubin in prolonged cholestasis when jaundice is overt), detection of bilirubin in urine via dipstick test is extremely helpful to confirm the presence of conjugated hyperbilirubinemia in a patient with mildly elevated direct fraction.

Several new techniques, although less convenient to perform, have added considerably to our understanding of bilirubin metabolism. First, studies using these methods demonstrate that, in normal persons or those with Gilbert's syndrome, almost 100% of the serum bilirubin is unconjugated; <3% is monoconjugated bilirubin. Second, in jaundiced patients with hepatobiliary disease, the total serum bilirubin concentration measured by these new, more accurate methods is lower than the values found with diazo methods. This finding suggests that there are diazo-positive compounds distinct from bilirubin in the serum of patients with hepatobiliary disease. Third, these studies indicate that, in jaundiced patients with hepatobiliary disease, monoglucuronides of bilirubin predominate over diglucuronides. Fourth, part of the direct-reacting bilirubin fraction includes conjugated bilirubin that is covalently linked to albumin. This albumin-linked fraction of conjugated bilirubin (*delta fraction*, *delta bilirubin*, or *biliprotein*) represents an important fraction of total serum bilirubin in patients with cholestasis and hepatobiliary disorders. The delta bilirubin is formed in serum when hepatic excretion of bilirubin glucuronides is impaired and the glucuronides accumulate in serum. By virtue of its tight binding to albumin, the clearance rate of delta bilirubin from serum approximates the half-life of albumin (12–14 days) rather than the short half-life of bilirubin (about 4 h).

The prolonged half-life of albumin-bound conjugated bilirubin accounts for two previously unexplained enigmas in jaundiced patients with liver disease: (1) that some patients with conjugated hyperbilirubinemia do not exhibit bilirubinuria during the recovery phase of their

disease because the delta bilirubin, although conjugated, is covalently bound to albumin and therefore not filtered by the renal glomeruli, and (2) that the elevated serum bilirubin level declines more slowly than expected in some patients who otherwise appear to be recovering satisfactorily. Late in the recovery phase of hepatobiliary disorders, all the conjugated bilirubin may be in the albumin-linked form.

MEASUREMENT OF URINE BILIRUBIN

Unconjugated bilirubin is always bound to albumin in the serum, is not filtered by the kidney, and is not found in the urine. Conjugated bilirubin is filtered at the glomerulus, and the majority is reabsorbed by the proximal tubules; a small fraction is excreted in the urine. Any bilirubin found in the urine is conjugated bilirubin. The presence of bilirubinuria on urine dipstick test (Ictotest) indicates an elevation of the conjugated bilirubin fraction that cannot be excreted from the liver and implies the presence of hepatobiliary disease. A false-negative result is possible in patients with prolonged cholestasis due to the predominance of delta bilirubin, which is covalently bound to albumin and therefore not filtered by the renal glomeruli.

APPROACH TO THE PATIENT

Jaundice

The goal of this chapter is not to provide an encyclopedic review of every condition that causes jaundice. Rather, the chapter is intended to offer a framework that helps a physician to evaluate the patient with jaundice in a logical way (Fig. 49-1).

The initial step is to perform appropriate blood tests in order to determine whether the patient has an isolated elevation of serum bilirubin. If so, is the bilirubin elevation due to an increased unconjugated or conjugated fraction? If the hyperbilirubinemia is accompanied by other liver test abnormalities, is the disorder hepatocellular or cholestatic? If cholestatic, is it intra- or extrahepatic? These questions can all be answered with a thoughtful history, physical examination, and interpretation of laboratory and radiologic tests and procedures.

The bilirubin present in serum represents a balance between input from the production of bilirubin and hepatic/biliary removal of the pigment. Hyperbilirubinemia may result from (1) overproduction of bilirubin; (2) impaired uptake, conjugation, or excretion of bilirubin; or (3) regurgitation of unconjugated or conjugated bilirubin from damaged hepatocytes or bile ducts. An increase in unconjugated bilirubin in serum results from overproduction, impaired uptake, or conjugation of bilirubin. An increase in conjugated bilirubin is due to decreased excretion into the bile ductules or backward leakage of the pigment. The initial steps in evaluating the patient with jaundice are to determine (1) whether the hyperbilirubinemia is predominantly conjugated or unconjugated in nature and (2) whether other biochemical liver tests are abnormal. The thoughtful interpretation of limited data permits a rational evaluation of the patient (Fig. 49-1). The following discussion will focus solely on the evaluation of the adult patient with jaundice.

ISOLATED ELEVATION OF SERUM BILIRUBIN

Unconjugated Hyperbilirubinemia The differential diagnosis of isolated unconjugated hyperbilirubinemia is limited (Table 49-1). The critical determination is whether the patient is suffering from a hemolytic process resulting in an overproduction of bilirubin (hemolytic disorders and ineffective erythropoiesis) or from impaired hepatic uptake/conjugation of bilirubin (drug effect or genetic disorders).

Hemolytic disorders that cause excessive heme production may be either inherited or acquired. Inherited disorders include spherocytosis, sickle cell anemia, thalassemia, and deficiency of red cell enzymes such as pyruvate kinase and glucose-6-phosphate dehydrogenase. In these conditions, the serum bilirubin level rarely exceeds 86 µmol/L (5 mg/dL). Higher levels may occur when there

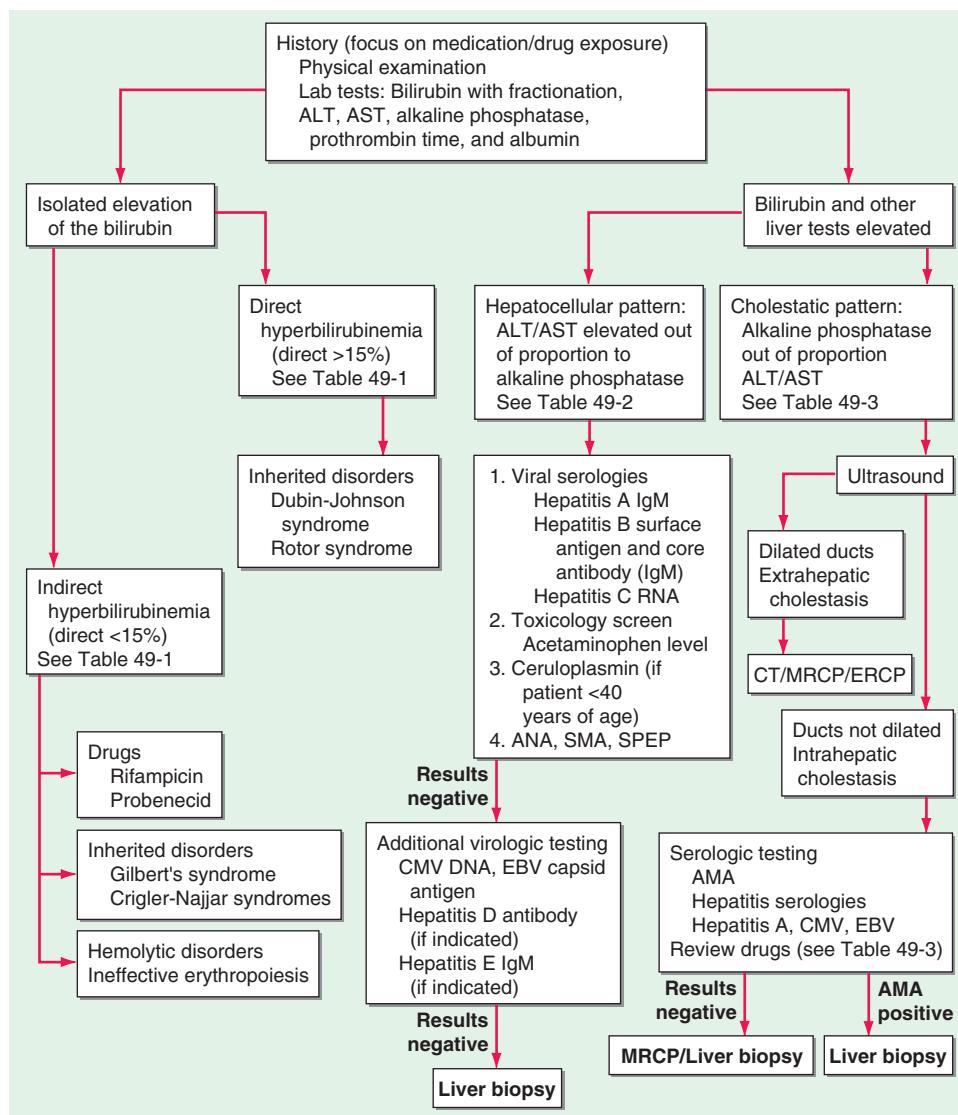


FIGURE 49-1 Evaluation of the patient with jaundice. ALT, alanine aminotransferase; AMA, antimitochondrial antibody; ANA, antinuclear antibody; AST, aspartate aminotransferase; CMV, cytomegalovirus; EBV, Epstein-Barr virus; ERCP, endoscopic retrograde cholangiopancreatography; LKM, liver-kidney microsomal antibody; MRCP, magnetic resonance cholangiopancreatography; SMA, smooth-muscle antibody; SPEP, serum protein electrophoresis.

TABLE 49-1 Causes of Isolated Hyperbilirubinemia

- I. Indirect hyperbilirubinemia
 - A. Hemolytic disorders
 - B. Ineffective erythropoiesis
 - C. Increased bilirubin production
 - 1. Massive blood transfusion
 - 2. Resorption of hematoma
 - D. Drugs
 - 1. Rifampin
 - 2. Probenecid
 - 3. Antibiotics—cephalosporins and penicillins
 - E. Inherited conditions
 - 1. Crigler-Najjar types I and II
 - 2. Gilbert's syndrome
- II. Direct hyperbilirubinemia (inherited conditions)
 - A. Dubin-Johnson syndrome
 - B. Rotor syndrome

is coexistent renal or hepatocellular dysfunction or in acute hemolysis, such as a sickle cell crisis. In evaluating jaundice in patients with chronic hemolysis, it is important to remember the high incidence of pigmented (calcium bilirubinate) gallstones found in these patients, which increases the likelihood of choledocholithiasis as an alternative explanation for hyperbilirubinemia.

Acquired hemolytic disorders include microangiopathic hemolytic anemia (e.g., hemolytic-uremic syndrome), paroxysmal nocturnal hemoglobinuria, spur cell anemia, immune hemolysis, and parasitic infections (e.g., malaria and babesiosis). Ineffective erythropoiesis occurs in cobalamin, folate, and iron deficiencies. Resorption of hematomas and massive blood transfusions both can result in increased hemoglobin release and overproduction of bilirubin.

In the absence of hemolysis, the physician should consider a problem with the hepatic uptake or conjugation of bilirubin. Certain drugs, including rifampin and probenecid, may cause unconjugated hyperbilirubinemia by diminishing hepatic uptake

of bilirubin. Impaired bilirubin conjugation occurs in three genetic conditions: Crigler-Najjar syndrome types I and II and Gilbert's syndrome. *Crigler-Najjar type I* is an exceptionally rare condition found in neonates and characterized by severe jaundice (bilirubin $>342 \mu\text{mol/L}$ [$>20 \text{ mg/dL}$]) and neurologic impairment due to kernicterus, frequently leading to death in infancy or childhood. These patients have a complete absence of bilirubin UDPGT activity; are totally unable to conjugate bilirubin; and hence cannot excrete it.

Crigler-Najjar type II is somewhat more common than type I. Patients live into adulthood with serum bilirubin levels of 103–428 $\mu\text{mol/L}$ (6–25 mg/dL). In these patients, mutations in the bilirubin UDPGT gene cause the reduction—typically $\leq 10\%$ —of the enzyme's activity. Bilirubin UDPGT activity can be induced by the administration of phenobarbital, which can reduce serum bilirubin levels in these patients. Despite marked jaundice, these patients usually survive into adulthood, although they may be susceptible to kernicterus under the stress of concurrent illness or surgery.

Gilbert's syndrome is also marked by the impaired conjugation of bilirubin due to reduced bilirubin UDPGT activity (typically 10–35% of normal). Patients with Gilbert's syndrome have mild unconjugated hyperbilirubinemia, with serum levels almost always $<103 \mu\text{mol/L}$ (6 mg/dL). The serum levels may fluctuate, and jaundice is often identified only during periods of stress, concurrent illness, alcohol use, or fasting. Unlike both Crigler-Najjar syndromes, Gilbert's syndrome is very common. The reported incidence is 3–7% of the population, with males predominating over females by a ratio of 1.5–7:1.

Conjugated Hyperbilirubinemia Elevated conjugated hyperbilirubinemia is found in two rare inherited conditions: *Dubin-Johnson syndrome* and *Rotor syndrome* (Table 49-1). Patients with either condition present with asymptomatic jaundice. The defect in Dubin-Johnson syndrome is the presence of mutations in the gene for MRP2. These patients have altered excretion of bilirubin into the bile ducts. Rotor syndrome may represent a deficiency of the major hepatic drug reuptake transporters OATP1B1 and OATP1B3. Differentiating between these syndromes is possible but is clinically unnecessary due to their benign nature.

ELEVATION OF SERUM BILIRUBIN WITH OTHER LIVER TEST ABNORMALITIES

The remainder of this chapter will focus on the evaluation of patients with conjugated hyperbilirubinemia in the setting of other liver test abnormalities. This group of patients can be divided into those with a primary hepatocellular process and those with intra- or extrahepatic cholestasis. This distinction, which is based on the history and physical examination as well as the pattern of liver test abnormalities, guides the clinician's evaluation (Fig. 49-1).

History A complete medical history is perhaps the single most important part of the evaluation of the patient with unexplained jaundice. Important considerations include the use of or exposure to any chemical or medication, whether physician-prescribed, over-the-counter, complementary, or alternative medicines (e.g., herbal and vitamin preparations) or other drugs such as anabolic steroids. The patient should be carefully questioned about possible parenteral exposures, including transfusions, intravenous and intranasal drug use, tattooing, and sexual activity. Other important points include recent travel history; exposure to people with jaundice; exposure to possibly contaminated foods; occupational exposure to hepatotoxins; alcohol consumption; the duration of jaundice; and the presence of any accompanying signs and symptoms, such as arthralgias, myalgias, rash, anorexia, weight loss, abdominal pain, fever, pruritis, and changes in the urine and stool. While none of the latter manifestations is specific for any one condition, any of them can suggest a diagnosis. A history of arthralgias and myalgias predating jaundice suggests hepatitis, either viral or drug related. Jaundice associated with the sudden onset of severe right-upper-quadrant

pain and shaking chills suggests choledocholithiasis and ascending cholangitis.

Physical Examination The general assessment should include evaluation of the patient's nutritional status. Temporal and proximal muscle wasting suggests long-standing disease such as pancreatic cancer or cirrhosis. Stigmata of chronic liver disease, including spider nevi, palmar erythema, gynecomastia, caput medusae, Dupuytren's contractures, parotid gland enlargement, and testicular atrophy, are commonly seen in advanced alcohol-related cirrhosis and occasionally in other types of cirrhosis. An enlarged left suprACLAVICULAR node (Virchow's node) or a periumbilical nodule (Sister Mary Joseph's nodule) suggests an abdominal malignancy. Jugular venous distention, a sign of right-sided heart failure, suggests hepatic congestion. Right pleural effusion even in the absence of clinically apparent ascites may be seen in advanced cirrhosis.

The abdominal examination should focus on the size and consistency of the liver, on whether the spleen is palpable and hence enlarged, and on whether ascites is present. Patients with cirrhosis may have an enlarged left lobe of the liver, which is felt below the xiphoid, and an enlarged spleen. A grossly enlarged nodular liver or an obvious abdominal mass suggests malignancy. An enlarged tender liver could signify viral or alcoholic hepatitis; an infiltrative process such as amyloidosis; or, less often, an acutely congested liver secondary to right-sided heart failure. Severe right-upper-quadrant tenderness with respiratory arrest on inspiration (Murphy's sign) suggests cholecystitis. Ascites in the presence of jaundice suggests either cirrhosis or malignancy with peritoneal spread.

Laboratory Tests A battery of tests are helpful in the initial evaluation of a patient with unexplained jaundice. These include total and direct serum bilirubin measurement with fractionation; determination of serum aminotransferase, alkaline phosphatase, and albumin concentrations; and prothrombin time tests. Enzyme tests (alanine aminotransferase [ALT], aspartate aminotransferase [AST], and alkaline phosphatase [ALP]) are helpful in differentiating between a hepatocellular process and a cholestatic process (Table 337-1; Fig. 49-1)—a critical step in determining what additional workup is indicated. Patients with a hepatocellular process generally have a rise in the aminotransferases that is disproportionate to that in ALP, whereas patients with a cholestatic process have a rise in ALP that is disproportionate to that of the aminotransferases. The serum bilirubin can be prominently elevated in both hepatocellular and cholestatic conditions and therefore is not necessarily helpful in differentiating between the two.

In addition to enzyme tests, all jaundiced patients should have additional blood tests—specifically, an albumin level and a prothrombin time—to assess liver function. A low albumin level suggests a chronic process such as cirrhosis or cancer. A normal albumin level is suggestive of a more acute process such as viral hepatitis or choledocholithiasis. An elevated prothrombin time indicates either vitamin K deficiency due to prolonged jaundice and malabsorption of vitamin K or significant hepatocellular dysfunction. The failure of the prothrombin time to correct with parenteral administration of vitamin K indicates severe hepatocellular injury.

The results of the bilirubin, enzyme, albumin, and prothrombin time tests will usually indicate whether a jaundiced patient has a hepatocellular or a cholestatic disease and offer some indication of the duration and severity of the disease. The causes and evaluations of hepatocellular and cholestatic diseases are quite different.

Hepatocellular Conditions Hepatocellular diseases that can cause jaundice include viral hepatitis, drug or environmental toxicity, alcohol, and end-stage cirrhosis from any cause (Table 49-2). Wilson's disease occurs primarily in young adults. Autoimmune hepatitis is typically seen in young to middle-aged women but may affect men and women of any age. Alcoholic hepatitis can be differentiated from viral and toxin-related hepatitis by the pattern of the aminotransferases: patients with alcoholic hepatitis typically have

TABLE 49-2 Hepatocellular Conditions That May Produce Jaundice

Viral hepatitis
Hepatitis A, B, C, D, and E
Epstein-Barr virus
Cytomegalovirus
Herpes simplex virus
Alcoholic hepatitis
Chronic liver disease and cirrhosis
Drug toxicity
Predictable, dose-dependent (e.g., acetaminophen)
Unpredictable, idiosyncratic (e.g., isoniazid)
Environmental toxins
Vinyl chloride
Jamaica bush tea—pyrrolizidine alkaloids
Kava kava
Wild mushrooms— <i>Amanita phalloides</i> , <i>A. verna</i>
Wilson's disease
Autoimmune hepatitis

an AST-to-ALT ratio of at least 2:1, and the AST level rarely exceeds 300 U/L. Patients with acute viral hepatitis and toxin-related injury severe enough to produce jaundice typically have aminotransferase levels >500 U/L, with the ALT greater than or equal to the AST. While ALT and AST values <8 times normal may be seen in either hepatocellular or cholestatic liver disease, values 25 times normal or higher are seen primarily in acute hepatocellular diseases. Patients with jaundice from cirrhosis can have normal or only slightly elevated aminotransferase levels.

When the clinician determines that a patient has a hepatocellular disease, appropriate testing for acute viral hepatitis includes a hepatitis A IgM antibody assay, a hepatitis B surface antigen and core IgM antibody assay, a hepatitis C viral RNA test, and, depending on the circumstances, a hepatitis E IgM antibody assay. The hepatitis C antibody can take up to 6 weeks to become detectable, making it an unreliable test if acute hepatitis C is suspected. Studies for hepatitis D, Epstein-Barr virus (EBV), and cytomegalovirus (CMV) may also be indicated. Ceruloplasmin is the initial screening test for Wilson's disease. Testing for autoimmune hepatitis usually includes antinuclear antibody and anti-smooth muscle antibody assays and measurement of specific immunoglobulins.

Drug-induced hepatocellular injury can be classified as either predictable or unpredictable. Predictable drug reactions are dose-dependent and affect all patients who ingest a toxic dose of the drug in question. The classic example is acetaminophen hepatotoxicity. Unpredictable or idiosyncratic drug reactions are not dose-dependent and occur in a minority of patients. A great number of drugs can cause idiosyncratic hepatic injury. Environmental toxins are also an important cause of hepatocellular injury. Examples include industrial chemicals such as vinyl chloride, herbal preparations containing pyrrolizidine alkaloids (Jamaica bush tea) or kava, and the mushrooms *Amanita phalloides* and *A. verna*, which contain highly hepatotoxic amatoxins.

Cholestatic Conditions When the pattern of the liver tests suggests a cholestatic disorder, the first step is to determine whether it is intra- or extrahepatic cholestasis (Fig. 49-1). Distinguishing intrahepatic from extrahepatic cholestasis may be difficult. History, physical examination, and laboratory tests often are not helpful. The next appropriate test is an ultrasound. The ultrasound is inexpensive, does not expose the patient to ionizing radiation, and can detect dilation of the intra- and extrahepatic biliary tree with a high degree of sensitivity and specificity. The absence of biliary dilation suggests intrahepatic cholestasis, while its presence indicates extrahepatic cholestasis. False-negative results occur in patients with

partial obstruction of the common bile duct or in patients with cirrhosis or primary sclerosing cholangitis (PSC), in which scarring prevents the intrahepatic ducts from dilating.

Although ultrasonography may indicate extrahepatic cholestasis, it rarely identifies the site or cause of obstruction. The distal common bile duct is a particularly difficult area to visualize by ultrasound because of overlying bowel gas. Appropriate next tests include computed tomography (CT), magnetic resonance cholangiopancreatography (MRCP), endoscopic retrograde cholangiopancreatography (ERCP), percutaneous transhepatic cholangiography (PTC), and endoscopic ultrasound (EUS). CT and MRCP are better than ultrasonography for assessing the head of the pancreas and for identifying choledocholithiasis in the distal common bile duct, particularly when the ducts are not dilated. ERCP is the "gold standard" for identifying choledocholithiasis. Beyond its diagnostic capabilities, ERCP allows therapeutic interventions, including the removal of common bile duct stones and the placement of stents. PTC can provide the same information as ERCP and it also allows for intervention in patients in whom ERCP is unsuccessful due to proximal biliary obstruction or altered gastrointestinal anatomy. MRCP has replaced ERCP as the initial diagnostic test in most cases. EUS displays sensitivity and specificity comparable to that of MRCP in the detection of bile duct obstruction and allows biopsy of suspected malignant lesions.

In patients with apparent *intrahepatic cholestasis*, the diagnosis is often made by serologic testing in combination with a liver biopsy. The list of possible causes of intrahepatic cholestasis is long and varied (**Table 49-3**). A number of conditions that typically cause a hepatocellular pattern of injury can also present as a cholestatic variant. Both hepatitis B and C viruses can cause cholestatic hepatitis (fibrosing cholestatic hepatitis). This disease variant has been reported in patients who have undergone solid organ transplantation. Hepatitis A and E, alcoholic hepatitis, and EBV or CMV infections may also present as cholestatic liver disease.

Drugs may cause intrahepatic cholestasis that is usually reversible after discontinuation of the offending agent, although it may take many months for cholestasis to resolve. Drugs most commonly associated with cholestasis are the anabolic and contraceptive steroids. Cholestatic hepatitis has been reported with chlorpromazine, imipramine, tolbutamide, sulindac, cimetidine, and erythromycin estolate. It also occurs in patients taking trimethoprim; sulfamethoxazole; and penicillin-based antibiotics such as ampicillin, dicloxacillin, and clavulanic acid. Rarely, cholestasis may be chronic and associated with progressive fibrosis despite early discontinuation of the offending drug. Chronic cholestasis has been associated with chlorpromazine and prochlorperazine.

Primary biliary cholangitis is an autoimmune disease predominantly affecting women and characterized by progressive destruction of interlobular bile ducts. The diagnosis is made by the detection of antimitochondrial antibody, which is found in 95% of patients. *Primary sclerosing cholangitis* is characterized by the destruction and fibrosis of larger bile ducts. The diagnosis of PSC is made with cholangiography (either MRCP or ERCP), which demonstrates the pathognomonic segmental strictures. Approximately 75% of patients with PSC also have inflammatory bowel disease.

The *vanishing bile duct syndrome* and *adult bile ductopenia* are rare conditions in which a decreased number of bile ducts are seen in liver biopsy specimens. This histologic picture is also seen in patients who develop chronic rejection after liver transplantation and in those who develop graft-versus-host disease after bone marrow transplantation. Vanishing bile duct syndrome also occurs in rare cases of sarcoidosis, in patients taking certain drugs (including chlorpromazine), and idiopathically.

There are also familial forms of intrahepatic cholestasis. The familial intrahepatic cholestatic syndromes include *progressive familial intrahepatic cholestasis* (PFIC) types 1–3 and *benign recurrent intrahepatic cholestasis* (BRIC) types 1 and 2. BRIC is characterized

TABLE 49-3 Cholestatic Conditions That May Produce Jaundice

I. Intrahepatic
A. Viral hepatitis
1. Fibrosing cholestatic hepatitis—hepatitis B and C
2. Hepatitis A, Epstein-Barr virus infection, cytomegalovirus infection
B. Alcoholic hepatitis
C. Drug toxicity
1. Pure cholestasis—anabolic and contraceptive steroids
2. Cholestatic hepatitis—chlorpromazine, erythromycin estolate
3. Chronic cholestasis—chlorpromazine and prochlorperazine
D. Primary biliary cholangitis
E. Primary sclerosing cholangitis
F. Vanishing bile duct syndrome
1. Chronic rejection of liver transplants
2. Sarcoidosis
3. Drugs
G. Congestive hepatopathy and ischemic hepatitis
H. Inherited conditions
1. Progressive familial intrahepatic cholestasis
2. Benign recurrent intrahepatic cholestasis
I. Cholestasis of pregnancy
J. Total parenteral nutrition
K. Nonhepatobiliary sepsis
L. Benign postoperative cholestasis
M. Paraneoplastic syndrome
N. Veno-occlusive disease
O. Graft-versus-host disease
P. Infiltrative disease
1. Tuberculosis
2. Lymphoma
3. Amyloidosis
Q. Infections
1. Malaria
2. Leptospirosis
II. Extrahepatic
A. Malignant
1. Cholangiocarcinoma
2. Pancreatic cancer
3. Gallbladder cancer
4. Ampullary cancer
5. Malignant involvement of the porta hepatis lymph nodes
B. Benign
1. Choledocholithiasis
2. Postoperative biliary strictures
3. Primary sclerosing cholangitis
4. Chronic pancreatitis
5. AIDS cholangiopathy
6. Mirizzi's syndrome
7. Parasitic disease (ascariasis)

by episodic attacks of pruritus, cholestasis, and jaundice beginning at any age, which can be debilitating but does not lead to chronic liver disease. Serum bile acids are elevated during episodes, but serum γ -glutamyltransferase (γ -GT) activity is normal. PFIC disorders begin at childhood and are progressive in nature. All three types of PFIC are associated with progressive cholestasis, elevated levels of serum bile acids, and similar phenotypes but different genetic mutations. Only type 3 PFIC is associated with high levels of γ -GT. *Cholestasis of pregnancy* occurs in the second and third trimesters and resolves after delivery. Its cause is unknown, but the

condition is probably inherited, and cholestasis can be triggered by estrogen administration.

Other causes of intrahepatic cholestasis include total parenteral nutrition (TPN); nonhepatobiliary sepsis; benign postoperative cholestasis; and a paraneoplastic syndrome associated with a number of different malignancies, including Hodgkin's disease, medullary thyroid cancer, renal cell cancer, renal sarcoma, T-cell lymphoma, prostate cancer, and several gastrointestinal malignancies. The term *Stauffer's syndrome* has been used for intrahepatic cholestasis specifically associated with renal cell cancer. In patients developing cholestasis in the intensive care unit, the major considerations should be sepsis, ischemic hepatitis ("shock liver"), and TPN-related jaundice. Jaundice occurring after bone marrow transplantation is most likely due to veno-occlusive disease or graft-versus-host disease. In addition to hemolysis, sickle cell disease may cause intrahepatic and extrahepatic cholestasis. Jaundice is a late finding in heart failure caused by hepatic congestion and hepatocellular hypoxia. Ischemic hepatitis is a distinct entity of acute hypoperfusion characterized by an acute and dramatic elevation in the serum aminotransferases followed by a gradual peak in serum bilirubin.

Jaundice with associated liver dysfunction can be seen in severe cases of *Plasmodium falciparum* malaria. The jaundice in these cases is due to a combination of indirect hyperbilirubinemia from hemolysis and both cholestatic and hepatocellular jaundice. Weil's disease, a severe presentation of leptospirosis, is marked by jaundice with renal failure, fever, headache, and muscle pain.

Causes of *extrahepatic cholestasis* can be split into malignant and benign (Table 49-3). Malignant causes include pancreatic, gallbladder, and ampullary cancers as well as cholangiocarcinoma. This last malignancy is most commonly associated with PSC and is exceptionally difficult to diagnose because its appearance is often identical to that of PSC. Pancreatic and gallbladder tumors as well as cholangiocarcinoma are rarely resectable and have poor prognoses. Ampullary carcinoma has the highest surgical cure rate of all the tumors that present as painless jaundice. Hilar lymphadenopathy due to metastases from other cancers may cause obstruction of the extrahepatic biliary tree.

Choledocholithiasis is the most common cause of extrahepatic cholestasis. The clinical presentation can range from mild right-upper-quadrant discomfort with only minimal elevations of enzyme test values to ascending cholangitis with jaundice, sepsis, and circulatory collapse. PSC may occur with clinically important strictures limited to the extrahepatic biliary tree. IgG4-associated cholangitis is marked by stricturing of the biliary tree. It is critical that the clinician differentiate this condition from PSC as it is responsive to glucocorticoid therapy. In rare instances, chronic pancreatitis causes strictures of the distal common bile duct, where it passes through the head of the pancreas. AIDS cholangiopathy is a condition that is usually due to infection of the bile duct epithelium with CMV or cryptosporidiosis and has a cholangiographic appearance similar to that of PSC. The affected patients usually present with greatly elevated serum alkaline phosphatase levels (mean, 800 IU/L), but the bilirubin level is often near normal. These patients do not typically present with jaundice.

■ GLOBAL CONSIDERATIONS

While extrahepatic biliary obstruction and drugs are common causes of new-onset jaundice in developed countries, infections remain the leading cause in developing countries. Liver involvement and jaundice are observed with numerous infections, particularly malaria, babesiosis, severe leptospirosis, infections due to *Mycobacterium tuberculosis* and the *Mycobacterium avium* complex, typhoid fever, infection with hepatitis viruses A–E, EBV, CMV, viral hemorrhagic fevers including Ebola virus, late phases of yellow fever, dengue fever, schistosomiasis, fascioliasis, clonorchiasis, opisthorchiasis, ascariasis, echinococcosis, hepatosplenic candidiasis, disseminated histoplasmosis, cryptococcosis, coccidioidomycosis, ehrlichiosis, chronic Q fever, yersiniosis,

brucellosis, syphilis, and leprosy. Bacterial infections that do not necessarily involve the liver and bile ducts may also lead to jaundice, as in cholestasis of sepsis. The presence of fever or abdominal pain suggests concurrent infection, sepsis, or complications from gallstones. The development of encephalopathy and coagulopathy in a jaundiced patient with no preexisting liver disease signifies acute liver failure, which warrants urgent liver transplant evaluation.

ACKNOWLEDGMENT

This chapter is a revised version of chapters that have appeared in prior editions of Harrison's in which Marshall M. Kaplan was a co-author with Daniel Pratt.

■ FURTHER READING

ERLINGER S et al: Inherited disorders of bilirubin transport and conjugation: New insights into molecular mechanisms and consequences. *Gastroenterology* 146:1625, 2014.

WOLKOFF AW et al: Bilirubin metabolism and jaundice, in *Schiff's Diseases of the Liver*, 11th ed, Schiff ER et al (eds). Oxford, UK, John Wiley & Sons, Ltd, 2012, pp 120–150.

coordination between diaphragmatic contraction and anterior abdominal wall relaxation, a response in some cases to intraluminal bowel stimuli; dietary alterations, manipulation of the intestinal microbiota, or biofeedback may be effective therapy. Occasionally, increased lumbar lordosis accounts for apparent abdominal distention.

Fat Weight gain with an increase in abdominal fat can result in an increase in abdominal girth and can be perceived as abdominal swelling. Abdominal fat may be caused by an imbalance between caloric intake and energy expenditure associated with a poor diet and sedentary lifestyle; it also can be a manifestation of certain diseases, such as Cushing's syndrome. Excess abdominal fat has been associated with an increased risk of insulin resistance and cardiovascular disease.

Fluid The accumulation of fluid within the abdominal cavity (ascites) often results in abdominal distention and is discussed in detail below. Grade 1 ascites is detectable only by ultrasonography; grade 2 ascites is detectable by physical examination; and grade 3 ascites results in marked abdominal distention.

Fetus Pregnancy results in increased abdominal girth. Typically, an increase in abdominal size is first noted at 12–14 weeks of gestation, when the uterus moves from the pelvis into the abdomen. Abdominal distention may be seen before this point as a result of fluid retention and relaxation of the abdominal muscles.

Feces In the setting of severe constipation or intestinal obstruction, increased stool in the colon leads to increased abdominal girth. These conditions are often accompanied by abdominal discomfort or pain, nausea, and vomiting and can be diagnosed by imaging studies.

Fatal Growth An abdominal mass can result in abdominal swelling. Neoplasms, abscesses, or cysts can grow to sizes that lead to increased abdominal girth. Enlargement of the intraabdominal organs, specifically the liver (hepatomegaly) or spleen (splenomegaly), or an abdominal aortic aneurysm can result in abdominal distention. Bladder distention also may result in abdominal swelling.

50

Abdominal Swelling and Ascites

Lawrence S. Friedman



ABDOMINAL SWELLING

Abdominal swelling is a manifestation of numerous diseases. Patients may complain of bloating or abdominal fullness and may note increasing abdominal girth on the basis of increased clothing or belt size. Abdominal discomfort is often reported, but pain is less frequent. When abdominal pain does accompany swelling, it is frequently the result of an intraabdominal infection, peritonitis, or pancreatitis. Patients with abdominal distension from *ascites* (fluid in the abdomen) may report the new onset of an inguinal or umbilical hernia. Dyspnea may result from pressure against the diaphragm and the inability to expand the lungs fully.

■ CAUSES

The causes of abdominal swelling can be remembered conveniently as the *six F*s: flatus, fat, fluid, fetus, feces, or a “fatal growth” (often a neoplasm).

Flatus Abdominal swelling may be the result of increased intestinal gas. The normal small intestine contains ~200 mL of gas made up of nitrogen, oxygen, carbon dioxide, hydrogen, and methane. Nitrogen and oxygen are consumed (swallowed), whereas carbon dioxide, hydrogen, and methane are produced intraluminally by bacterial fermentation. Increased intestinal gas can occur in a number of conditions. *Aerophagia*, the swallowing of air, can result in increased amounts of oxygen and nitrogen in the small intestine and lead to abdominal swelling. Aerophagia typically results from gulping food; chewing gum; smoking; or as a response to anxiety, which can lead to repetitive belching. In some cases, increased intestinal gas is the consequence of bacterial metabolism of excess fermentable substances such as lactose and other oligosaccharides, which can lead to production of hydrogen, carbon dioxide, or methane. In many cases, the precise cause of abdominal distention cannot be determined. In some persons, particularly those with irritable bowel syndrome and bloating, the subjective sense of abdominal pressure is attributable to impaired intestinal transit of gas rather than increased gas volume. Abdominal distention—an objective increase in girth—is the result of a lack of

APPROACH TO THE PATIENT

Abdominal Swelling

HISTORY

Determining the etiology of abdominal swelling begins with history-taking and a physical examination. Patients should be questioned regarding symptoms suggestive of malignancy, including weight loss, night sweats, and anorexia. Inability to pass stool or flatus together with nausea or vomiting suggests bowel obstruction, severe constipation, or an ileus (lack of peristalsis). Increased eructation and flatus may point toward aerophagia or increased intestinal production of gas. Patients should be questioned about risk factors for or symptoms of chronic liver disease, including excessive alcohol use and jaundice, which suggest ascites. Patients should also be asked about symptoms of other medical conditions, including heart failure and tuberculosis, which may cause ascites.

PHYSICAL EXAMINATION

Physical examination should include an assessment for signs of systemic disease. The presence of lymphadenopathy, especially supraclavicular lymphadenopathy (*Virchow's node*), suggests metastatic abdominal malignancy. Care should be taken during the cardiac examination to evaluate for elevation of jugular venous pressure (JVP); *Kussmaul's sign* (elevation of the JVP during inspiration); a pericardial knock, which may be seen in heart failure or constrictive pericarditis; or a murmur of tricuspid regurgitation. Spider angiomas, palmar erythema, dilated superficial veins around the umbilicus (*caput medusae*), and gynecomastia suggest liver disease.

The abdominal examination should begin with inspection for the presence of uneven distention or an obvious mass. Auscultation should follow. The absence of bowel sounds or the presence

of high-pitched localized bowel sounds points toward an ileus or intestinal obstruction. An umbilical venous hum may suggest the presence of portal hypertension, and a harsh bruit over the liver is heard rarely in patients with hepatocellular carcinoma or alcohol-associated hepatitis. Abdominal swelling caused by intestinal gas can be differentiated from swelling caused by fluid or a solid mass by percussion; an abdomen filled with gas is tympanic, whereas an abdomen containing a mass or fluid is dull to percussion. The absence of abdominal dullness, however, does not exclude ascites, because a minimum of 1500 mL of ascitic fluid is required for detection on physical examination. Finally, the abdomen should be palpated to assess for tenderness, a mass, enlargement of the spleen or liver, or presence of a nodular liver suggesting cirrhosis or tumor. Light palpation of the liver may detect pulsations suggesting retrograde vascular flow from the heart in patients with right-sided heart failure, particularly tricuspid regurgitation.

IMAGING AND LABORATORY EVALUATION

Abdominal x-rays can be used to detect dilated loops of bowel suggesting intestinal obstruction or ileus. Abdominal ultrasonography can detect as little as 100 mL of ascitic fluid, hepatosplenomegaly, a nodular liver, or a mass. Ultrasonography is often inadequate to detect retroperitoneal lymphadenopathy or a pancreatic lesion because of overlying bowel gas. If malignancy or pancreatic disease is suspected, CT can be performed. CT may also detect changes associated with advanced cirrhosis and portal hypertension (Fig. 50-1).

Laboratory evaluation should include liver biochemical testing, serum albumin level measurement, and prothrombin time determination (international normalized ratio) to assess hepatic function as well as a complete blood count to evaluate for the presence of cytopenias that may result from portal hypertension or of leukocytosis, anemia, and thrombocytosis that may result from systemic infection. Serum amylase and lipase levels should be checked to evaluate the patient for acute pancreatitis. Urinary protein quantitation is indicated when nephrotic syndrome, which may cause ascites, is suspected. Hydrogen and methane absorbed from the intestine are not metabolized by the host and are excreted in expired air, and detection of increased amounts of these gases in expired breath is the basis for tests used to

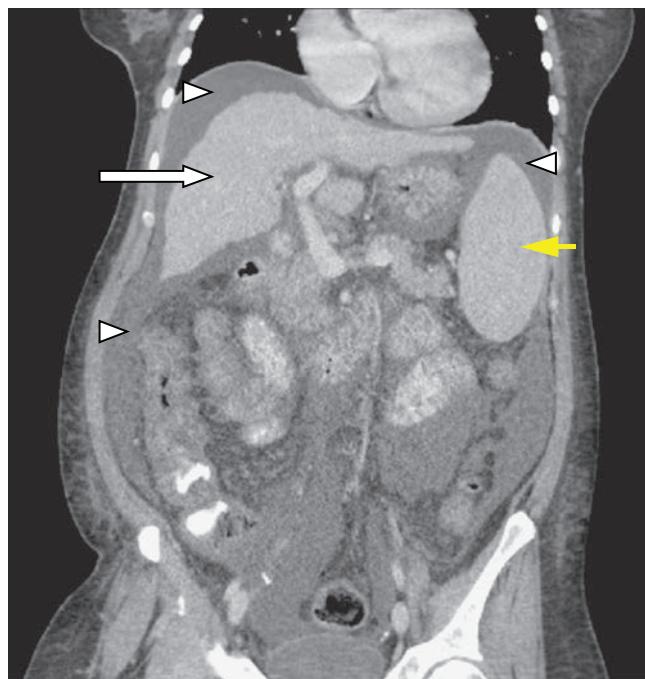


FIGURE 50-1 CT of a patient with a cirrhotic, nodular liver (white arrow), splenomegaly (yellow arrow), and ascites (arrowheads).

diagnose carbohydrate (e.g., lactose) malabsorption and small intestinal bacterial overgrowth.

In selected cases, the hepatic venous pressure gradient (pressure across the liver between the portal and hepatic veins) can be measured via cannulation of the hepatic vein to confirm that ascites is caused by cirrhosis (**Chap. 344**). In some cases, a liver biopsy may be necessary to confirm cirrhosis.

ASCITES

PATHOGENESIS IN THE PRESENCE OF CIRRHOsis

Ascites in patients with cirrhosis is the result of portal hypertension and renal salt and water retention. Similar mechanisms contribute to ascites formation in heart failure. Portal hypertension signifies elevation of the pressure within the portal vein. According to Ohm's law, pressure is the product of resistance and flow. Increased hepatic resistance occurs by several mechanisms. First, the development of hepatic fibrosis, which defines cirrhosis, disrupts the normal architecture of the hepatic sinusoids and impedes normal blood flow through the liver. Second, activation of hepatic stellate cells, which mediate fibrogenesis, leads to smooth-muscle contraction and fibrosis. Finally, cirrhosis is associated with a decrease in endothelial nitric oxide synthetase (eNOS) production, which results in decreased nitric oxide production and increased intrahepatic vasoconstriction.

The development of cirrhosis is also associated with increased systemic circulating levels of nitric oxide (in contrast to the decrease seen intrahepatically), as well as increased levels of vascular endothelial growth factor and tumor necrosis factor, that result in splanchnic arterial vasodilation. Vasodilation of the splanchnic circulation results in pooling of blood and a decrease in the effective circulating volume, which is perceived by the kidneys as hypovolemia. Compensatory vasoconstriction via release of antidiuretic hormone ensues; the consequences are free water retention and activation of the sympathetic nervous system and the renin-angiotensin-aldosterone system, which lead in turn to renal sodium and water retention.

PATHOGENESIS IN THE ABSENCE OF CIRRHOsis

Ascites in the absence of cirrhosis generally results from peritoneal carcinomatosis, peritoneal infection, or pancreatic disease. Peritoneal carcinomatosis can result from primary peritoneal malignancies such as mesothelioma or sarcoma, abdominal malignancies such as gastric or colonic adenocarcinoma, or metastatic disease from breast or lung carcinoma or melanoma (Fig. 50-2). The tumor cells lining the

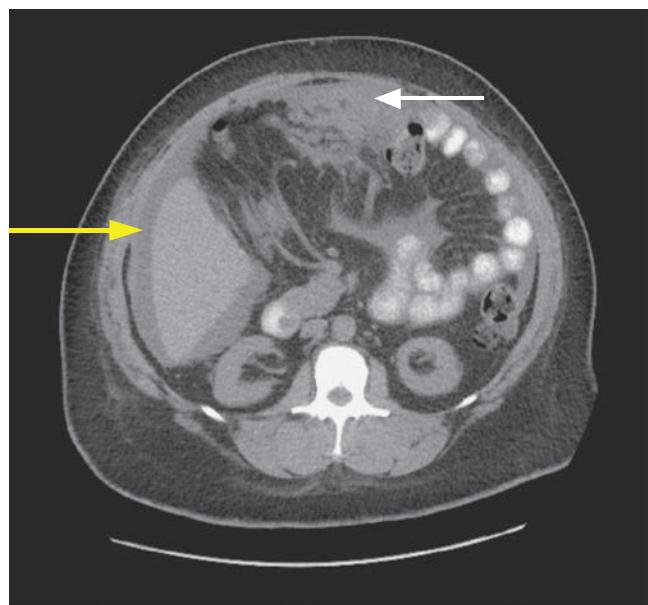


FIGURE 50-2 CT of a patient with peritoneal carcinomatosis (white arrow) and ascites (yellow arrow).

peritoneum produce a protein-rich fluid that contributes to the development of ascites. Fluid from the extracellular space is drawn into the peritoneum, further contributing to the development of ascites. Tuberculous peritonitis causes ascites via a similar mechanism; tubercles deposited on the peritoneum exude a proteinaceous fluid. Pancreatic ascites results from leakage of pancreatic enzymes into the peritoneum.

■ CAUSES

Cirrhosis accounts for 84% of cases of ascites. Cardiac ascites, peritoneal carcinomatosis, and “mixed” ascites resulting from cirrhosis and a second disease account for 10–15% of cases. Less common causes of ascites include massive hepatic metastasis, infection (tuberculosis, *Chlamydia* infection), pancreatitis, and renal disease (nephrotic syndrome). Rare causes of ascites include hypothyroidism and familial Mediterranean fever.

■ EVALUATION

Once the presence of ascites has been confirmed, the etiology of the ascites is best determined by *paracentesis*, a bedside procedure in which a needle or small catheter is passed transcutaneously to extract ascitic fluid from the peritoneum. The lower quadrants are the most frequent sites for paracentesis. The left lower quadrant is preferred because of the greater depth of ascites and the thinner abdominal wall. Paracentesis is a safe procedure even in patients with coagulopathy; complications, including abdominal wall hematomas, hypotension, hepatorenal syndrome, and infection, are infrequent.

Once ascitic fluid has been extracted, its gross appearance should be examined. Turbid fluid can result from the presence of infection or tumor cells. White, milky fluid indicates a triglyceride level >200 mg/dL (and often >1000 mg/dL), which is the hallmark of *chylous ascites*. Chylous ascites results from lymphatic disruption that may occur with trauma, cirrhosis, tumor, tuberculosis, or certain congenital abnormalities. Dark brown fluid can reflect a high bilirubin concentration and indicates biliary tract perforation. Black fluid may indicate the presence of pancreatic necrosis or metastatic melanoma.

The ascitic fluid should be sent for measurement of albumin and total protein levels, cell and differential counts, and, if infection is suspected, Gram's stain and culture, with inoculation into blood culture bottles at the patient's bedside to maximize the yield. A serum albumin level should be measured simultaneously to permit calculation of the *serum-ascites albumin gradient* (SAAG).

The SAAG is useful for distinguishing ascites caused by portal hypertension from nonportal hypertensive ascites (Fig. 50-3). The SAAG reflects the pressure within the hepatic sinusoids and correlates with the hepatic venous pressure gradient. The SAAG is calculated by subtracting the ascitic albumin concentration from the serum albumin level and does not change with diuresis. A SAAG ≥ 1.1 g/dL reflects the presence of portal hypertension and indicates that the ascites is due to increased pressure in the hepatic sinusoids. According to Starling's law, a high SAAG reflects the oncotic pressure that counterbalances the portal pressure. Possible causes include cirrhosis, cardiac ascites,

hepatic vein thrombosis (Budd-Chiari syndrome), sinusoidal obstruction syndrome (veno-occlusive disease), or massive liver metastases. A SAAG <1.1 g/dL indicates that the ascites is not related to portal hypertension, as in tuberculous peritonitis, peritoneal carcinomatosis, or pancreatic ascites.

For high-SAAG (≥ 1.1) ascites, the ascitic protein level can provide further clues to the etiology (Fig. 50-3). An ascitic protein level of ≥ 2.5 g/dL indicates that the hepatic sinusoids are normal and are allowing passage of protein into the ascites, as occurs in cardiac ascites, early Budd-Chiari syndrome, or sinusoidal obstruction syndrome. An ascitic protein level <2.5 g/dL indicates that the hepatic sinusoids have been damaged and scarred and no longer allow passage of protein, as occurs with cirrhosis, late Budd-Chiari syndrome, or massive liver metastases. Pro-brain-type natriuretic peptide (BNP) is a natriuretic hormone released by the heart as a result of increased volume and ventricular wall stretch. High levels of BNP in serum occur in heart failure and may be useful in identifying heart failure as the cause of high-SAAG ascites.

Further tests are indicated only in specific clinical circumstances. When secondary peritonitis resulting from a perforated hollow viscus is suspected, ascitic glucose and lactate dehydrogenase (LDH) levels can be measured. In contrast to “spontaneous” bacterial peritonitis, which may complicate cirrhotic ascites (see “Complications,” below), secondary peritonitis is suggested by an ascitic glucose level <50 mg/dL, an ascitic LDH level higher than the serum LDH level, and the detection of multiple pathogens on ascitic fluid culture. When pancreatic ascites is suspected, the ascitic amylase level should be measured and is typically >1000 mg/dL. Cytology can be useful in the diagnosis of peritoneal carcinomatosis. At least 50 mL of fluid should be obtained and sent for immediate processing. Tuberculous peritonitis is typically associated with ascitic fluid lymphocytosis but can be difficult to diagnose by paracentesis. A smear for acid-fast bacilli has a diagnostic sensitivity of only 0–3%; a culture increases the sensitivity to 35–50%. In patients without cirrhosis, an elevated ascitic adenosine deaminase level has a sensitivity of $>90\%$ for tuberculous ascites when a cut-off value of 30–45 U/L is used. When the cause of ascites remains uncertain, laparotomy or laparoscopy with peritoneal biopsies for histology and culture remains the gold standard.

TREATMENT

Ascites

The initial treatment for cirrhotic ascites is restriction of sodium intake to 2 g/d. When sodium restriction alone is inadequate to control ascites, oral diuretics—typically the combination of spironolactone and furosemide—are used to increase urinary sodium excretion. Spironolactone is an aldosterone antagonist that inhibits sodium resorption in the distal convoluted tubule of the kidney. Use of spironolactone may be limited by hyponatremia, hyperkalemia, and painful gynecomastia. If the gynecomastia is

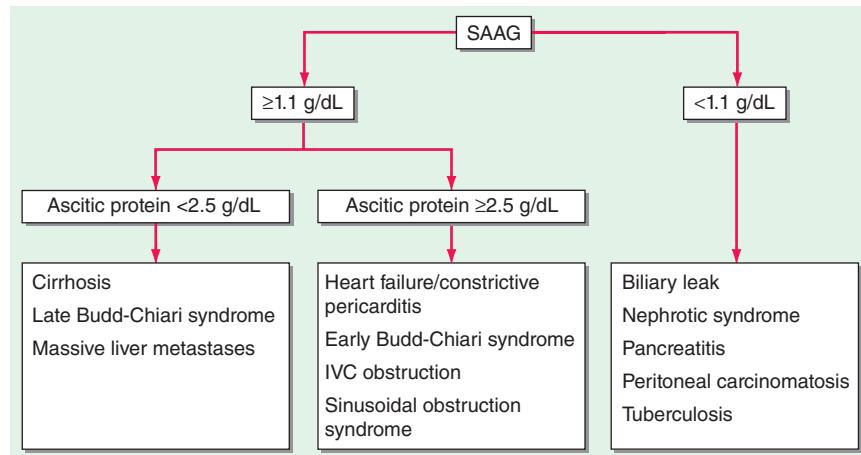


FIGURE 50-3 Algorithm for the diagnosis of ascites according to the serum-ascites albumin gradient (SAAG). IVC, inferior vena cava.

distressing, amiloride (5–40 mg/d) may be substituted for spironolactone. Furosemide is a loop diuretic that is generally combined with spironolactone in a ratio of 40:100; maximal daily doses of spironolactone and furosemide are 400 mg and 160 mg, respectively. Fluid intake may be restricted in patients with hyponatremia.

Refractory cirrhotic ascites is defined by the persistence of ascites despite sodium restriction and maximal (or maximally tolerated) diuretic use. Pharmacologic therapy for refractory ascites includes the addition of midodrine, an α_1 -adrenergic agonist, or clonidine, an α_2 -adrenergic agonist, to diuretic therapy. These agents act as vasoconstrictors, counteracting splanchnic vasodilation. Midodrine alone or in combination with clonidine improves systemic hemodynamics and control of ascites over that obtained with diuretics alone. Although β -adrenergic blocking agents (beta blockers) are often prescribed to prevent variceal hemorrhage in patients with cirrhosis, the use of beta blockers in patients with refractory ascites may be associated with decreased survival rates.

When medical therapy alone is insufficient, refractory cirrhotic ascites can be managed by repeated large-volume paracentesis (LVP) or a transjugular intrahepatic peritoneal shunt (TIPS)—a radiologically placed portosystemic shunt that decompresses the hepatic sinusoids. Intravenous (IV) infusion of albumin accompanying LVP decreases the risk of “postparacentesis circulatory dysfunction” and death. Patients undergoing LVP should receive IV albumin infusions of 6–8 g/L of ascitic fluid removed. TIPS placement is superior to LVP in reducing the reaccumulation of ascites but is associated with an increased frequency of hepatic encephalopathy, with no difference in mortality rates. The Alfapump system, which consists of an automated pump and tunneled peritoneal catheter that transports ascites from the peritoneal cavity to the urinary bladder, has shown promise in the management of refractory ascites but is associated with a higher frequency of technical difficulties and renal dysfunction.

Malignant ascites does not respond to sodium restriction or diuretics. Patients must undergo serial LVPs, transcutaneous drainage catheter placement, or, rarely, creation of a peritoneovenous shunt (a shunt from the abdominal cavity to the vena cava) or placement of the Alfapump system, if available.

Ascites caused by tuberculous peritonitis is treated with standard antituberculosis therapy. Noncirrhotic ascites of other causes is treated by correction of the precipitating condition.

COMPLICATIONS

Spontaneous bacterial peritonitis (SBP; [Chap. 132](#)) is a common and potentially lethal complication of cirrhotic ascites. Occasionally, SBP also complicates ascites caused by nephrotic syndrome, heart failure, acute hepatitis, and acute liver failure but is rare in malignant ascites. Patients with SBP generally note an increase in abdominal girth; however, abdominal tenderness is found in only 40% of patients, and rebound tenderness is uncommon. Patients may present with fever, nausea, vomiting, or the new onset or an exacerbation of preexisting hepatic encephalopathy.

In hospitalized patients with ascites, paracentesis within 12 hours of admission reduces mortality because of early detection of SBP. SBP is defined by a polymorphonuclear neutrophil (PMN) count of $\geq 250/\mu\text{L}$ in the ascitic fluid. Cultures of ascitic fluid should be performed in blood culture bottles and typically reveal one bacterial pathogen. The presence of multiple pathogens in the setting of an elevated ascitic PMN count suggests *secondary peritonitis* from a ruptured viscus or abscess ([Chap. 132](#)). The presence of multiple pathogens without an elevated PMN count suggests bowel perforation from the paracentesis needle. SBP is generally the result of enteric bacteria that have translocated across an edematous bowel wall. The most common pathogens are gram-negative rods, including *Escherichia coli* and *Klebsiella*, as well as streptococci and enterococci.

Treatment of SBP with an antibiotic such as IV cefotaxime is generally effective against gram-negative and gram-positive aerobes. A

5-day course of treatment is sufficient if the patient improves clinically. Nosocomial or health care-acquired SBP is frequently caused by multi-drug-resistant bacteria, and initial antibiotic therapy should be guided by the local bacterial epidemiology.

Cirrhotic patients with a history of SBP, an ascitic fluid total protein concentration $<1\text{ g/dL}$, or active gastrointestinal bleeding should receive prophylactic antibiotics to prevent SBP; oral daily ciprofloxacin or, where available, norfloxacin is commonly used. IV ceftriaxone may be used in hospitalized patients. Diuresis increases the activity of ascitic fluid protein opsonins and may decrease the risk of SBP.

Hepatic hydrothorax occurs when ascites, often caused by cirrhosis, migrates via fenestrae in the diaphragm into the pleural space. This condition can result in shortness of breath, hypoxia, and infection. Treatment is similar to that for cirrhotic ascites and includes sodium restriction, diuretics, and, if needed, thoracentesis or TIPS placement. Chest tube placement should be avoided.

ACKNOWLEDGMENT

The author thanks Dr. Kathleen E. Corey for contributions to this chapter in prior editions of the textbook.

FURTHER READING

- ADEBAYO D et al: Refractory ascites in liver cirrhosis. *Am J Gastroenterol* 114:40, 2019.
- BARBA E et al: Correction of abdominal distention by biofeedback-guided control of abdominothoracic muscular activity in a randomized, placebo-controlled trial. *Clin Gastroenterol Hepatol* 15:1922, 2017.
- BERNARDI M et al: Albumin infusion in patients undergoing large-volume paracentesis: A meta-analysis of randomized trials. *Hepatology* 55:1172, 2012.
- EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER: EASL Clinical Practice Guidelines for the management of patients with decompensated cirrhosis. *J Hepatol* 69:406, 2018.
- FARIAS AQ et al: Serum B-type natriuretic peptide in the initial workup of patients with new onset ascites: A diagnostic accuracy study. *Hepatology* 59:1043, 2014.
- FERNANDEZ J et al: Prevalence and risk factors of infections by multiresistant bacteria in cirrhosis: A prospective study. *Hepatology* 55:1551, 2012.
- GE PS, RUNYON BA: Role of plasma BNP in patients with ascites: Advantages and pitfalls. *Hepatology* 59:751, 2014.
- JOHN S, FRIEDMAN LS: Portal hypertensive ascites: Current status. *Curr Hepatol Rep* 19:226, 2020.
- JOHN S, THULUVATH PJ: Hyponatremia in cirrhosis: Pathophysiology and management. *World J Gastroenterol* 21:3197, 2015.
- LIZAOLA B et al: Review article: the diagnostic approach and current management of chylous ascites. *Aliment Pharmacol Ther* 46:816, 2017.
- MALAGELADA JR et al: Bloating and abdominal distension: Old misconceptions and current knowledge. *Am J Gastroenterol* 112:1221, 2017.
- ORMAN ES et al: Paracentesis is associated with reduced mortality in patients hospitalized with cirrhosis and ascites. *Clin Gastroenterol Hepatol* 12:496, 2014.
- RUNYON BA: Introduction to the revised American Association for the Study of Liver Diseases Practice Guideline management of adult patients with ascites due to cirrhosis 2012. *Hepatology* 57:165, 2013.
- RUNYON BA et al: The serum-ascites albumin gradient is superior to the exudate-transudate concept in the differential diagnosis of ascites. *Ann Intern Med* 117:215, 1992.
- SORT P et al: Effect of intravenous albumin on renal impairment and mortality in patients with cirrhosis and spontaneous bacterial peritonitis. *N Engl J Med* 341:403, 1999.
- WILLIAMS JW Jr, SIMEL DL: The rational clinical examination. Does this patient have ascites? How to divine fluid in the abdomen. *JAMA* 267:2645, 1992.

Section 7 Alterations in Renal and Urinary Tract Function

51

Interstitial Cystitis/Bladder Pain Syndrome

R. Christopher Doiron, J. Curtis Nickel



DEFINITION

A condition associated with bladder inflammation and pain, with what were thought to be discrete bladder ulcerations, was first described in 1887. The description of the classic bladder-wall ulcer—now referred to as a *Hunner lesion*—became known as *interstitial cystitis* (IC). The first generally accepted definition of IC was derived from a National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) consensus of experts in the field in 1998. The NIDDK criteria used to define IC included typical cystoscopic findings such as glomerulations (submucosal petechial hemorrhages of the urothelium) or Hunner lesions. However, over time, the syndrome experienced by patients, including bladder and/or pelvic pain with associated urinary storage symptoms of urinary frequency and urgency, negative urine cultures, and no specific identifiable causes, became known as *interstitial cystitis/bladder pain syndrome* (IC/BPS).

The nomenclature and definitions have evolved, but the contemporary definitions accepted by the American Urological Association, the Canadian Urological Association, the International Continence Society, the Society for Urodynamics and Female Urology, and the European Society for the Study of IC/BPS, although they all differ somewhat in language and specifics, generally reflect several fundamental concepts common in the disease: (1) it is chronic in nature; (2) it causes pain perceived to be attributable to the bladder; (3) this pain occurs in the presence of lower urinary tract symptoms (LUTS); and (4) pain outside the bladder—in the pelvis, perineum, genitals, abdomen, and beyond—is common.

The following definition incorporates the major descriptions by all international groups interested in the diagnosis and management of IC/BPS: an unpleasant sensation (pain, pressure, discomfort) perceived to be related to the urinary bladder, associated with LUTS of >6 weeks' duration, in the absence of infection or other identifiable causes.

A generalized urologic chronic pelvic pain syndrome (UCPPS) is referenced in the literature and is thought to encompass two distinct urologic chronic pain disorders: IC/BPS, which may be present in men and women, and chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS), which is present only in men. The latter refers to a urologic pain disorder with pain localized to the perineum and/or male genitals, with or without LUTS. IC/BPS can exist independent of CP/CPPS in men. In reality, the urologic chronic pain disorders often have overlapping symptom presentations and may share common etiologic and pathophysiologic origins, but the focus of the current chapter will be on IC/BPS.

ETIOLOGY AND PATHOGENESIS

Pinning a single etiology to a diagnosis of IC/BPS has been an endeavor fraught with uncertainty that ultimately has failed thus far. Instead, it is much more likely and widely accepted that IC/BPS represents a syndrome or constellation of interrelated disease processes that manifest in a spectrum of disease that reaches beyond the bladder. While the search for a single etiology soldiers on, we will review here a collection of proposed theories.

■ INFECTION AND THE URINARY MICROBIOTA

Bacterial infection of the urothelium has long been regarded as a major suspect in the etiology of IC/BPS but has never been definitively shown to cause the disease. It is not uncommon for patients presenting with IC/BPS to describe a long history of “urinary tract infections” (UTIs); these patients often have undergone multiple courses of treatment with one or multiple antibiotics prescribed by their physicians. Often, however, in patients with IC/BPS, the benefit of antibiotic treatment is short-lived, urine culture results are negative, and the return of symptoms is inevitable.

Although studies examining the role of microbiologic organisms in this patient population are numerous and the results conflicting, far more studies have yielded negative results rather than positive findings. Furthermore, our understanding of the urinary microbiota continues to expand, rendering older studies using outdated and insensitive cultivation techniques less relevant.

Using state-of-the-art, culture-independent techniques for microorganism identification, investigators observe subtle differences between the urinary microbiota of IC/BPS patients and that of healthy controls and between IC/BPS patients experiencing symptom flares and IC/BPS patients not in flare. The clinical relevance of these findings is still not fully understood. As the study of the urinary microbiota continues to unfold, researchers and clinicians believe that although a single causative microbe is unlikely, dysbiosis or disturbance in the microbial ecology of the lower urinary tract may be responsible for flares or symptom patterns experienced by IC/BPS patients.

■ AUTOIMMUNITY

The consideration of IC/BPS as a disorder of the immune system stems from the observation of a significant prevalence of autoimmune disorders in IC/BPS patients; several historical studies have identified anti-urothelial antibodies within the bladder mucosa of IC/BPS patients. Furthermore, although IC/BPS is not a pathologic diagnosis, there are widely accepted, recognizable patterns of inflammatory infiltration in the bladder mucosa of this patient population, including lymphoplasmacytic infiltrates, stromal edema and fibrosis, urothelial denudation, and detrusor mastocytosis. Thus, although it is likely that immune disturbances cause the condition in a subset of patients (for example, in those with associated Sjögren's syndrome), researchers and clinicians have been unable to leverage this knowledge into a clear description, and its clinical relevance is not fully understood.

■ INFLAMMATION

It is well established that a subset of patients suffering from IC/BPS clearly have associated bladder inflammation of unknown etiology. The best described of these patients are those with Hunner lesions—discrete inflammatory lesions, previously believed to be ulcers, that have a well-characterized inflammatory profile on histologic and pathologic analysis. While Hunner lesions are easily identified under direct vision by cystoscopy, a spectrum of other, less obvious inflammatory patterns in the bladder is associated with infiltration of acute and chronic inflammatory cells and mast cells. This inflammation observed on histologic analysis can be so subtle that it cannot be recognized under direct visual examination of the bladder with cystoscopy.

Investigators in the Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP) Research Network have found that, among patients with UCPPS, women exhibit more robust inflammatory responses to stimulation of Toll-like receptor 2 (TLR2) and TLR4. Furthermore, an increased response to stimulation of TLR4 predicts more severe symptoms, widespread pain (vs pelvic/bladder pain only), and a higher number of chronic overlapping pain conditions (COPCs). Further studies aimed at a better understanding of these findings are under way.

■ UROTHELIAL DYSFUNCTION

Urothelial Permeability and the Glycosaminoglycan Layer The stratified epithelium of the bladder—the urothelium—is composed of basal precursor cells, intermediate cells, and a layer of

specialized, superficial epithelial cells called *umbrella cells*. Collectively, these layers are responsible for the various functions of the bladder lining. One important function of the urothelium is to provide a robust barrier layer. This function is fulfilled by the dense layering of glycosaminoglycans (GAGs) on the luminal surface of the urothelium along with a complex arrangement of numerous intercellular tight junctions among urothelial cells that protects the underlying bladder interstitium from the constituents of the urine resting in the bladder.

Defects in this barrier function—either disruptions in the GAG layer or disruptions in the epithelial layer itself or its cellular junctions—have been proposed as a possible mechanism for bladder pain in IC/BPS patients. This theory, while still popular, lacks definitive evidence supporting it as the etiology of this disease.

Antiproliferative Factor The discovery that urothelial cells from IC/BPS patients appear to grow far more slowly than urothelial cells from a healthy control population led to the identification of antiproliferative factor (APF). Although APF initially showed promise as a sensitive and specific urine biomarker for IC/BPS, this idea has not been widely adopted, and the etiologic role of APF is not yet fully understood.

■ PELVIC ORGAN CROSSTALK

The observation of dysfunction and symptoms in multiple organ systems, including gastrointestinal, gynecologic, and genital organs, in patients with IC/BPS is so common that it might be considered the norm. Mechanisms of neural sensitization in patients with chronic pain have been reported, and abnormalities in the autonomic nervous system have been observed among IC/BPS patients. Again, although these observations apply in a subset of patients, their broader application to the heterogeneous IC/BPS patient population as a clear cause of disease is not warranted.

■ NEUROBIOLOGIC CONTRIBUTIONS AND CENTRAL SENSITIZATION

One breakthrough by the MAPP Research Network is an investigation of the role of structural and functional alterations in the brains of patients with UCPPS. The network's innovative methods of correlating clinical and deep phenotyping data with functional MRI data identified such structural and functional differences. These differences were later shown to successfully predict the progression of symptoms in a cohort of 52 patients with UCPPS. Although the relevant study did not differentiate between IC/BPS and CP/CPPS patients, the findings are nevertheless informative, and further longitudinal studies are ongoing.

In addition to the novel MAPP-led findings using neuroimaging, quantitative sensory testing (QST) methods have been used to investigate the sensory processing mechanisms in UCPPS patients. The findings—generalized pain hypersensitivity and altered endogenous inhibitory pain control systems among UCPPS patients—further support a hypothesis of a central sensitization phenotype in urologic chronic pelvic pain. The clinical implications of observed neural alterations and multisensory hypersensitivity remain under investigation.

Although a single etiology for this clinically heterogeneous pain syndrome may never be identified, efforts to do so have revealed much about its pathogenesis in subsets of patients and have provided valuable insight into specific patient phenotypes. The challenge for researchers and clinicians moving forward will be to unify clinical phenotypes with these proposed underlying mechanisms of disease and to integrate this knowledge into clinically actionable interventions that may provide meaningful outcomes.

EPIDEMIOLOGY

The prevalence of IC/BPS has been difficult to determine because definitions and diagnostic criteria (in the absence of a definitive diagnostic test or biomarker) are constantly evolving. In addition, the various methods used in attempting to describe the syndrome's epidemiology (patient self-reports, symptom-based surveys, physician visits, population-based databases) have been problematic and have made comparisons of results challenging. Many studies have historically been performed only in female populations. Currently, it is estimated that

2.7–6.5% of North American women experience symptoms consistent with a diagnosis of IC/BPS. Fewer than 10% of women who experience these symptoms actually have a diagnosis of IC/BPS. The syndrome does occur in men, with a reported 10:1 female-to-male ratio, but it is thought that the condition is dramatically underreported in men.

Some predictors of the development of IC/BPS have been suggested through an analysis of retrospective observational studies of childhood disorders and adverse childhood experiences (ACEs), including childhood UTI, childhood bowel and bladder dysfunction, and childhood sexual trauma. Furthermore, it has been well established that IC/BPS patients exhibit a remarkable prevalence of COPCs such as fibromyalgia, irritable bowel syndrome (IBS), chronic back pain, and chronic fatigue syndrome (CFS). Recent MAPP-led studies have shown that more than one-third of IC/BPS patients have one COPC (IBS, fibromyalgia, or CFS), while up to 10% have multiple COPCs. Thus, these conditions might be considered risk factors for the development of IC/BPS.

CLINICAL MANIFESTATIONS

Patients with IC/BPS, both female and male, present with varying degrees of discomfort and/or pain perceived to be related to the bladder and associated with urinary storage symptoms, including daytime and nighttime urinary frequency and urinary urgency. For some patients, urinary symptoms (the most common complaint after bladder pain) are the most bothersome, while for most patients, bladder pain causes the most distress and most significantly affects quality of life. Unfortunately, the majority of patients with IC/BPS present with both types of symptoms, as patients void frequently to relieve pain (or because of fear of bladder pain). Typically, this combination of bladder pain and urinary frequency severely impacts patients' quality of life, social interactions, and physical activities.

Pain-mapping studies have been used to identify different pain phenotypes within the disease. Nickel and colleagues first described a bladder-only phenotype present in 20% of a cohort of female IC/BPS patients, whereas up to 80% of patients described pain in the pelvis and at least one site beyond. Common associated conditions include IBS (40%), pelvic floor dysfunctional pain syndrome (40–60%), vulvodynia (17%), fibromyalgia (36%), CFS (10%), and chronic back pain (47%). As described above, these multiorgan symptoms may be due to central nervous system sensitization and associated spinal crosstalk, which may promote phenotypic progression as patients with one pain syndrome slowly progress to another. Subsequent MAPP-led studies among a more heterogeneous UCPPS cohort of men and women have supported this concept of specific pain phenotypes, reporting a pelvic-pain-only phenotype in 25% of participants and pain in the pelvis and beyond in up to 75%.

Another important finding from the MAPP investigations is their identification of not just a pelvic-pain-only phenotype but also of a bladder-focused phenotype. The latter phenotype was identified by patients' responses to two RAND Interstitial Cystitis Epidemiology (RICE) survey questions: whether they had “painful bladder filling” and/or “painful urinary urgency.” Most female UCPPS patients (88%) responded “yes” to at least one of these questions. The bladder-focused phenotype was associated with more severe urologic symptoms and worse quality of life.

Patients present with unique pain trajectories. Some initially have mild discomfort that progresses over many years to pain with bladder filling and finally to chronic unremitting pelvic pain with only short periods of relief with urination. Other patients begin with UTI-like symptoms and acute bladder and urethral pain with urinary frequency and urgency; these manifestations persist as a chronic cystitis-like syndrome despite negative cultures and no benefit from antimicrobial therapy. Still other patients report a waxing and waning of pain over time, with flares exacerbated by diet, anxiety/stress, infection, or hormone cycle (typically with increased pain prior to menses). In a longitudinal study of UCPPS patients followed over a 12-month period during routine care for their disease, MAPP investigators described 60% of patients' symptoms as stable, 20% as improved, and 20% as worsened.

TABLE 51-1 Workup of Patients by a Primary Care Practitioner or General Internist

STEPS IN WORKUP	SPECIFICS
History/physical examination	Conduct a pelvic exam (recommended). Categorize symptoms as bladder/pelvis focused and/or extending beyond the pelvis.
Urinalysis	Perform a urine culture. If the culture is positive, conduct sensitivity testing.
Consideration of patient-centered treatment options if satisfied with diagnosis ^a	Begin with conservative measures. Introduce further symptom-specific treatments as needed.
Referral to an appropriate specialist under certain conditions	Referral should follow if: <ul style="list-style-type: none"> the diagnosis is unclear microscopic or gross hematuria is present the condition is refractory to treatment symptoms are severe the presentation is complex

^aSee text.

APPROACH TO THE PATIENT

Interstitial Cystitis/Bladder Pain Syndrome

Patients with IC/BPS present to their family physician or internist with pelvic pain that typically increases in severity with bladder filling, other associated pain, and various degrees of urinary symptomatology. The course that should be followed by primary care practitioners or general internists during the patient's workup and before referral to a specialist is outlined in **Table 51-1**. Most of these physicians will not move beyond a suspected diagnosis and conservative advice; that is acceptable. Patients with IC/BPS can often represent diagnostic challenges, and referral to an appropriate subspecialist is warranted if any diagnostic uncertainty remains.

A diagnosis of IC/BPS is often missed and delayed for many years because physicians tend to silo patients into various medical-specialty streams on the basis of the predominant or most bothersome symptom. For example, patients presenting with pelvic pain in which flares are associated with monthly menstrual cycles may be referred to gynecologists. Patients with abdominal/pelvic pain associated with diarrhea and/or constipation tend to be referred to gastroenterologists, while those with generalized muscle and joint pain, perhaps associated with fatigue, are referred to rheumatologists. Patients with urinary symptoms and bladder pain are treated for UTIs (even with negative urine cultures) or overactive bladder—a common bladder condition associated with urinary frequency and urgency, but not pain.

It would be simple if the approach to patients presenting with pelvic pain was only to determine the actual pelvic organ and/or disease causing the symptoms. However, spinal crosstalk, phenotype progression over time, central sensitization, and COPCs complicate the picture. Since only ~20–25% of patients eventually diagnosed with IC/BPS have bladder-only disease, one must not be bladder-centric in approach but rather must consider the entire patient. The provider must determine the patient's "clinical picture"—that is, the patient's unique presenting clinical phenotype.

Urologists have adapted a system of clinical symptom categorization for patients with UC/PPS. UPOINT, which includes documenting the contribution of six distinct domains—Urinary, Psychosocial, Organ-specific, Infection, Neurologic, and Tenderness (as in pelvic floor muscle tenderness)—has helped categorize patient symptoms and allows the practitioner to focus their management on the most bothersome domain, while helping to avoid neglecting domains that are often forgotten. While used by many urologists managing this condition, UPOINT is not as effective in IC/BPS as it is in male CP/PPS, probably because all IC/BPS patients would be categorized, by definition, in the U and O domains.

A further simplified clinical approach to the assessment of patients with symptoms of IC/BPS is to classify patients with perceived bladder pain (a mandatory criterion for diagnosis) into one of two categories: (1) a "pelvic-pain-only" category, which would include the "bladder-pain-only," pelvic floor dysfunctional pain, and associated gynecologic pain groups; or (2) a "pelvic pain and beyond" category, which would include patients with associated COPCs (such as IBS and fibromyalgia). This approach has been supported by recent observations from the MAPP investigators.

The contribution of psychosocial parameters, such as depression, catastrophizing, anxiety, and stress, and their impact on pain and disability cannot be overlooked and are important to ascertain in all cases. This approach to clinical phenotyping will let the physician tailor a unique treatment plan for each individual patient, using combinations of local bladder, pelvic floor, or more general systemic therapies.

DIAGNOSIS

IC/BPS is a clinically heterogeneous condition whose lack of a clear etiopathogenesis presents difficulties in diagnosis. In making a diagnosis of exclusion, clinicians must rule out other confusable diseases and identify to the best of their ability the phenotypic presentation of the presenting patient. Although attempts have been made to establish a set of diagnostic criteria in the past, the specified criteria have proven overly stringent and too exclusive to be clinically useful. Furthermore, although several guidelines exist to aide in decision making in diagnostic investigations, most investigations serve merely to rule out other pathology. In contrast, history and physical examination, along with some simple laboratory testing, are the most reliable tools with which to establish a diagnosis of IC/BPS. Details of relevant investigations, some of which may be beyond the scope of the general practitioner or internist, are presented here. Table 51-1 offers an approach for the general practitioner, and **Table 51-2** provides a more complete summary of diagnostic recommendations.

HISTORY AND PHYSICAL (INCLUDING FREQUENCY/VOLUME CHARTS)

A thorough history and physical examination are of utmost importance in diagnosing IC/BPS. A history of the patient's pain symptoms is a logical place to start. The nature, intensity, and timing of the pain are all significant factors. Some patients will be less explicit than others in describing their pain and may instead describe a sense of pressure, burning, or vague fullness in the pelvis or bladder area.

All aspects of the patient's pain should be explored, as many patients' pain will not be limited to the pelvis or bladder but will be associated with the genitals, anus or rectum, perineum, abdomen, and beyond. Furthermore, although pain is commonly experienced with bladder filling, patients may also have suprapubic tenderness or pressure with voiding or burning or pain in the bladder, urethra, or perineum, with radiation into the vagina for women or the prostate, penis, and testicles

TABLE 51-2 Recommendations for Investigations in Patients with Suspected Interstitial Cystitis/Bladder Pain Syndrome

MANDATORY	RECOMMENDED	OPTIONAL	NOT RECOMMENDED
History	Frequency/volume chart	Ultrasound/pelvic imaging	Potassium sensitivity test
Physical examination	Urinalysis	Postvoid residual	Urodynamics
	Urine culture	Urine cytology	Bladder biopsy
	Symptom scores	Intravesical anesthetic bladder challenge	
	Cystoscopy	Hydrodistension	

Source: Adapted from A Cox et al: CUA guideline: Diagnosis and treatment of interstitial cystitis/bladder pain syndrome. Can Urol Assoc J 10:E136, 2016.

for men. In male patients, distinguishing IC/BPS from CP/CPPS can be challenging. Physicians must assess for more widespread pain locations outside the pelvis; screening for COPCs, particularly IBS, fibromyalgia, CFS, back pain, and headache, is important in adequately addressing the clinical impact of IC/BPS.

Eliciting and understanding associated LUTS—specifically urinary frequency, urgency, and nocturia—should be another focus of the history. While several confusable diseases can present with LUTS, the manifestation of IC/BPS as voiding dysfunction can help guide treatment decisions and is often a significant focus of bother for the patient. Having patients complete frequency/volume charts, noting the time and volume of each urination over a 24-hour period, can help provide objective evidence of a LUTS history and facilitate follow-up during and after treatment.

Physical examination should focus on the abdomen, pelvis, genitals, and pelvic floor. The degree of pelvic floor relaxation (i.e., degree of muscle tension and/or spasm) during examination is important to note. Trigger points in the pelvic floor musculature and any areas of localized spasticity should be identified. In women, an examination of the vulva, vaginal mucosa, and urethral meatus is essential to identify the presence of vulvodynia (vulvar mucosal pain with no identifiable cause) or any signs of genitourinary syndrome of menopause. In men, an examination of the external genitalia and a digital rectal (prostate) examination as well as a similar pelvic floor examination should be included to rule out related pathology.

■ SYMPTOM SCORES

The quality of a history can be elevated by an accompanying validated, objective measurement of the patient's symptoms. Although several relevant tools exist, the Interstitial Cystitis Symptom Index (ICSI) and the Interstitial Cystitis Problem Index (ICPI) are the most widely used, are commonly employed in research trials as outcome measures, and are straightforward enough for the practitioner to perform in an outpatient setting. These short questionnaires document pain severity, urinary frequency, urgency, and nocturia as well as the bother experienced from each of these symptoms.

More recently, MAPP investigators have suggested that pain and urinary symptoms should be assessed independently using two separate questionnaires: the Genitourinary Pain Index (GUPI) to assess pain and the ICSI to separately assess urinary symptoms. This suggestion is based on their finding of variable effects of urologic pain versus urinary symptoms on quality of life and mental health. Although symptom scores should not be relied on as diagnostic tools, their utility in establishing objective baseline measures to monitor response to treatment and symptoms over time can be valuable to the patient and the practitioner.

■ URINE STUDIES

Urine studies (urinalysis, culture, sensitivity, and cytology) should be included in the workup of a patient in whom IC/BPS is suspected. However, their role is mostly in ruling out other confusable disease rather than in aiding in the diagnosis of IC/BPS. A microscopic examination of the urine can reveal abnormalities attributable to the kidney that may warrant referral to a nephrologist; microscopic hematuria may trigger cystoscopic examination and referral to a urologist. The presenting symptoms of IC/BPS often mimic those of UTI, which must be ruled out by urine cultures. It is important to recognize that IC/BPS patients are subject to at least as great a risk of UTIs as the general population and that UTI should be considered when a flare in symptoms is reported. Finally, urine cytology should be considered if a diagnosis of bladder cancer is suspected or if there is a history of hematuria.

■ IMAGING, CYSTOSCOPY, AND URODYNAMICS

More intensive investigations and imaging studies should be considered in specific scenarios but need not be routinely performed. Abdominal and pelvic imaging studies in selected patients can help identify anatomic abnormalities of the upper or lower urinary tracts, diagnose urolithiasis or masses in the upper urinary tract, and rule out hydronephrosis, which may suggest obstructive uropathy. Furthermore,

brain imaging with functional MRI and quantitative sensory testing may prove beneficial in establishing a central sensitization phenotype; however, this is an emerging field of investigation, and routine brain imaging and sensory testing are currently not recommended.

Cystoscopy is used to rule out bladder pathology—most importantly, bladder cancer. Moreover, cystoscopy plays an important role in phenotyping IC/BPS and is required for identification of Hunner lesions. Although a broad consensus is lacking, the authors and others advocate for routine cystoscopic evaluation when IC/BPS is suspected, given the potential therapeutic implications and the ability of this measure to make phenotype-directed therapies possible. Finally, urodynamics testing should be reserved for specific scenarios—for example, cases in which complex voiding dysfunction may be contributing to the presentation.

■ INTRAVESICAL ANESTHETIC BLADDER CHALLENGE AND HYDRODISTENSION

An intravesical anesthetic bladder challenge (using intravesical lidocaine) can be done in the outpatient setting and can help distinguish bladder-focused pain from pelvic pain of other causes. It can further be harnessed as a therapeutic strategy if the patient experiences an improvement in symptoms. Similarly, hydrodistension, which requires a general or regional anesthetic, can play a diagnostic or therapeutic role. Bladder capacities of <400 mL under general anesthesia have correlated with worse pain and poor prognosis. The diagnostic role of post-hydrodistension inspection for bladder glomerulations has been suggested as a possible important clinical differentiation, although the utility of identifying and grading glomerulations is debated.

TREATMENT

Clinical Phenotyping

The UPOINT phenotyping tool introduced in 2009 was the first clinical tool to recognize that patients presenting with pelvic pain syndromes are a heterogeneous population with disease of unclear etiology that makes it difficult to predict outcomes in individuals with standard therapies. UPOINT is based on a patient-centered approach: individualized treatments are matched to patient evaluations by phenotyping of patients using six distinct clinical domains—urinary, psychosocial, organ-specific, infectious, neurologic, and tenderness. Since its initial publication, follow-up phenotyping studies have indicated that UPOINT is likely better than other methods in establishing phenotypic pain patterns in the clinic setting as local (bladder specific or pelvic pain only) or widespread (pelvic pain and beyond). Similarly, identifying inflammatory subtypes (e.g., Hunner lesion patients) and psychological parameters can help organize a patient's management plan and make it more likely that interventions will be successful. Applying an individualized multimodal treatment approach has proven beneficial in clinical practice. A collection of treatment options that might be directed at different domains of disease is presented in Fig. 51-1.

Although many of these treatments would be considered outside the scope of a general practitioner or general internist, it is important for the practitioner to be aware of them. In general, treatment should begin with more conservative measures, moving on to oral regimens or more invasive procedures if the patient's condition does not improve. A patient-centered approach is paramount in considering treatment escalation. The American Urological Association's IC/BPS guidelines provide a measure of overall efficacy of each individual therapy and a suggested order of implementation (tiered approach), but, because of the inability to predict the response to specific therapies, it is more clinically pragmatic to choose a multimodal approach based on the individual patient's presenting clinical phenotype or "clinical picture." Physicians are better positioned to implement this approach than are surgeons (urologists and gynecologists), who tend to be more organ- and surgery-focused when treating IC/BPS patients.

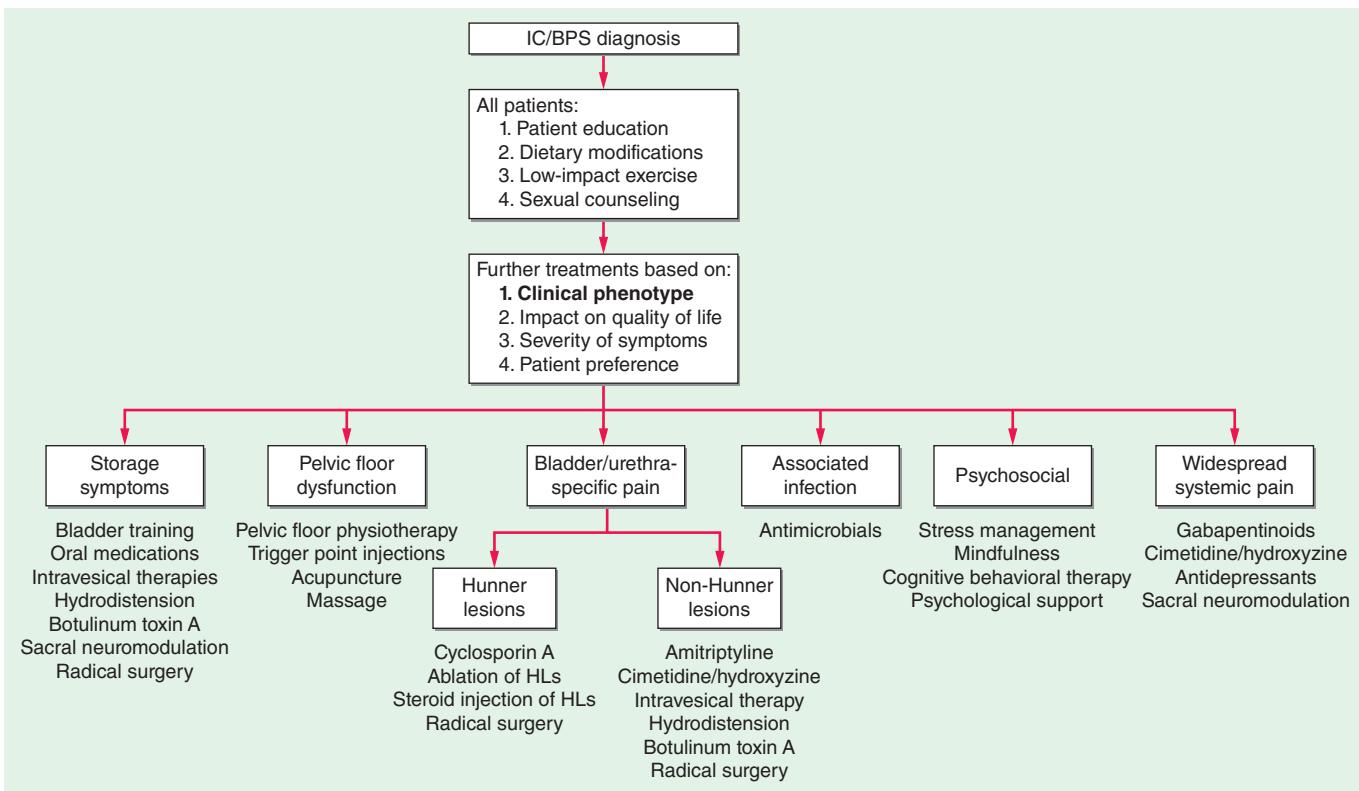


FIGURE 51-1 Proposed management paradigm for the treatment of interstitial cystitis (IC)/bladder pain syndrome (BPS). HLs, Hunner lesions.

■ CONSERVATIVE MEASURES

Conservative measures should be implemented for all patients with a diagnosis of IC/BPS. These therapies tend to be simple and inexpensive to introduce, pose little risk of significant side effects, and can be intensified or abandoned on the basis of the patient's response.

Patient Education Patient education and empowerment are paramount in this chronic pain disorder. Patients have often seen multiple practitioners prior to their diagnosis of IC/BPS. Acknowledging their suffering while educating them about their disease can go a long way in terms of relieving stress and anxiety related to an unknown and poorly understood problem. This acknowledgement also helps to develop a therapeutic patient–provider relationship. Setting realistic expectations and understanding that cure is not the goal constitute an important first step. Several resources are available for patients to explore at their own leisure.

Dietary Modifications Although limited evidence supports the role of dietary modifications, it has long been recognized that certain foods can trigger flares in IC/BPS patients and that simple dietary modifications can result in meaningful improvements in symptoms. Common dietary triggers include acidic and spicy foods and/or drinks, caffeinated or alcoholic beverages, artificial sweeteners, and/or gluten products; this list is by no means exhaustive, and dietary modifications should be made on an individual basis.

Pelvic Floor Physiotherapy Involvement of the pelvic floor in the pain syndrome can be ascertained on physical examination. Randomized studies have shown that, for patients who are found to have dysfunctional pelvic floors—muscle spasm, trigger points, or tenderness—contributing to their pain syndrome, pelvic floor physiotherapy may be beneficial. The musculoskeletal anatomy of the pelvic floor is complex; finding a provider with training specifically on the pelvic floor can be difficult but is crucial. Because accessing this resource may be financially burdensome for the patient, working together to find a way to obtain this helpful adjunctive therapy is important.

Psychological Interventions Mental health and psychosocial factors have long been identified as significantly prevalent in the IC/BPS

population and can impact disease and quality-of-life outcomes. There is some indication that, in IC/BPS and other related chronic pain conditions, mindfulness and cognitive behavioral therapy may improve outcomes. Challenges in accessing these therapies are a major barrier, and there is a general lack of consensus on which specific interventions are best suited to individual patients.

■ MEDICAL THERAPIES

Only two medications are currently approved by the U.S. Food and Drug Administration (FDA) for the treatment of IC/BPS: pentosan polysulfate sodium (PPS) given orally and dimethyl sulfoxide (DMSO) given intravesically. However, a collection of medications, administered orally or intravesically, are commonly used (albeit off-label) for this purpose.

Oral Therapies • PPS The only FDA-approved oral medication for IC/BPS has recently come under scrutiny because of reports regarding its association with vision-threatening maculopathy. Although causation has yet to be established, given its marginal benefit in the treatment of IC/BPS, the authors recommend against the long-term use of this medication. For patients currently taking PPS, the risks and benefits of treatment must be weighed. Consideration of a trial of weaning off the medication may be in the best interest of the patient. Any patients experiencing vision-related complaints while taking PPS should undergo immediate ophthalmologic assessment.

ANTIBIOTICS IC/BPS is not an infectious condition, and thus, antibiotics should have no role in treatment. Furthermore, the overwhelming majority of IC/BPS patients will have received at least one course, if not several courses, of antibiotics at some point in the course of their disease. Nevertheless, it is not unreasonable to administer a single course of antibiotics (after obtaining a sample for urine culture and sensitivity testing) if the patient has never previously received such therapy.

AMITRIPTYLINE Amitriptyline's pharmacologic activity is attributable primarily to its anticholinergic properties, its serotonin and norepinephrine uptake-inhibiting activity, and its sedative effects, which may include an antihistaminic pathway. Amitriptyline has been used to treat IC/BPS and other chronic pain syndromes. Studies support

the use of amitriptyline in IC/BPS patients while recognizing that the benefits can be marginal and associated with significant side effects.

CIMETIDINE AND HYDROXYZINE Early nonrandomized studies of hydroxyzine in the treatment of IC/BPS yielded promising results. As with amitriptyline, hydroxyzine's mechanism of action in treating IC/BPS is not fully understood and is likely multifactorial, owing largely to its antihistaminic effect via H₁-receptor antagonism but perhaps also its anticholinergic properties, its anxiolytic and sedative effects, and its inhibition of mast cell secretion and activation. An underpowered randomized study found no significant difference in symptom improvement between hydroxyzine and placebo.

After the modest success reported for hydroxyzine, cimetidine—an H₂-receptor antagonist—was investigated as another possible IC/BPS treatment. Only two observational studies and a single, small randomized clinical trial were completed and showed improvement in suprapubic pain and LUTS, particularly urinary frequency.

GABAPENTINOIDs Although no randomized studies of gabapentinoids have been performed in the IC/BPS population, these agents have been shown to improve symptoms in related chronic pain conditions. Furthermore, observational studies have shown some efficacy in IC/BPS. In properly selected patients in whom neuropathic pain is suspected, this medication class may have some success.

CYCLOSPORINA Despite its significant side effect profile, cyclosporin A has been investigated as treatment for IC/BPS refractory to other, more standard therapies. Because of its potent anti-inflammatory properties (it is used extensively in organ transplant recipients), this drug is particularly effective in IC/BPS patients with Hunner lesions, although the improvement in symptoms is modest. Side effects, including hypertension and nephrotoxicity, must be carefully monitored for, and the medication is typically reserved for patients in whom standard therapies have failed.

Intravesical Therapies Intravesical instillations remain a mainstay in the management of bladder-specific pain. Although this treatment modality typically requires an office visit, able and motivated patients can be trained to administer the medication at home. Patients' responses are variable, and the treatment is not curative, but it can significantly change the trajectory of disease in some patients and can rescue those experiencing symptom flares. Intravesical instillations can be administered as induction therapy; maintenance strategies have been proposed and can be effective for properly selected patients. A plethora of agents—most of them used off-label for this indication—have been investigated. The best-studied options are reviewed here.

DMSO DMSO, a solvent with anti-inflammatory properties, has been used in intravesical treatment for IC/BPS for several years. Despite a lack of high-quality evidence (with efficacy documented in only one placebo-controlled randomized clinical trial), DMSO remains the only FDA-approved intravesical medication for IC/BPS treatment. Its use has fallen out of favor, however, largely because of its unpleasant side effect of halitosis (its elimination via the lungs is associated with a garlic-like odor). Although the degree of improvement in symptoms is highly variable (60–95%), DMSO remains in the armamentarium of intravesical therapies.

HEPARIN Heparin, a glycosaminoglycan, was first investigated as a treatment for IC/BPS in light of the glycosaminoglycan layer deficiency theory of IC/BPS etiology; in animal models, heparin was shown to restore areas of damaged urothelium. Although there are no randomized clinical trials showing its efficacy, several observational studies have suggested benefit. Furthermore, in current practice, heparin is commonly administered with other medications as part of an intravesical "cocktail." Systemic absorption is minimal and appears not to affect coagulation parameters.

LIDOCAINe Lidocaine is commonly used as a local anesthetic and has been investigated as an option for intravesical treatment for IC/BPS. This agent works by blocking sensory nerves in the urothelium. Its

absorption and efficacy increase by alkalinization, commonly through coadministration with sodium bicarbonate.

HYALURONIC ACID AND CHONDROITIN SULFATE Hyaluronic acid and, more recently, chondroitin sulfate have been targeted as potential intravesical therapies because of their potentially restorative impact on the glycosaminoglycan layer. As is the case with most intravesical therapies, the quality of evidence is low, but there appears to be a modest benefit, with few side effects. Thus, these agents remain as options for patients whose disease is refractory to more standard therapies.

Trigger Point Injection Injection of a local anesthetic into myofascial trigger points in the pelvic floor (identified on physical examination) is a minimally invasive, practical therapy that can be administered during an office visit and can provide relief in properly selected patients. As with all therapies for this chronic pain condition, patient selection is paramount. The evidence supporting this treatment is largely anecdotal and based on expert opinion. A small nonblinded observational trial found a 72% success rate among women diagnosed with chronic pelvic pain and trigger points on physical examination; 33% of women were completely pain free after the injection. Although robust prospective trials are needed, this modality adds to the clinician's armamentarium in the treatment of IC/BPS.

SURGICAL THERAPIES

Treatment of Hunner Lesions A unique subset of IC/BPS patients have Hunner lesions. Direct treatment of these lesions through ablation with cauterization or laser treatment or, alternatively, injection of the lesion with a glucocorticoid improves symptoms in 70–90% of these patients. Hunner lesions tend to be recurrent, however; thus, treated patients still require follow-up, with consideration of a multimodal approach to their disease.

Hydrodistension Recent reports confirm that one of the oldest therapies for IC/BPS, hydrodistension under general anesthesia, provides some benefit in up to 54% of patients. Short- and long-term adverse effects (including bladder perforation and long-term bladder wall fibrosis) and the temporary nature of the benefit (with symptoms typically recurring within 3–12 months) mean that repeated bladder distension may not be an ideal long-term management strategy.

Onabotulinum Toxin A Onabotulinum toxin A (i.e., Botox) has been used to treat IC/BPS. There has, however, never been a randomized, placebo-controlled study evaluating onabotulinum toxin A injection into the detrusor muscle as monotherapy in this disease. Several randomized studies have evaluated this agent's efficacy and have shown improvements in symptoms; unfortunately, these trials often use onabotulinum toxin A in combination with hydrodistension, lack a placebo arm, and do not control for LUTS. Attributing benefit to this treatment is thus challenging. Furthermore, the side effect of acute urinary retention can be catastrophic in IC/BPS patients, whose pain is often secondary to bladder filling.

Sacral Neuromodulation In properly selected patients, sacral neuromodulation (SNM) can provide symptom relief in IC/BPS. Its use for this purpose is off-label, and it is not FDA approved for pain therapy. Nevertheless, SNM is FDA approved for treatment of bladder overactivity—a common symptom in IC/BPS patients—that is refractory to standard therapies. Although studies evaluating improvements in pain have shown variable results, a recent meta-analysis examining 17 observational trials (but no randomized clinical trials) of the use of SNM for IC/BPS does support its efficacy, with a statistically significant pooled treatment success rate of up to 84%. Side effects related to SNM must be considered, and the procedure carries with it a high rate of revision surgery, which is needed in up to half of patients undergoing this treatment.

Radical Surgery Radical surgery for the treatment of IC/BPS is reserved as a modality of desperation for the most refractory patients.

Options range from substitution cystoplasty to cystectomy with urinary diversion. Although improved symptoms and quality of life may result, such surgery is a potentially morbid operation and patient selection must be very specific.

COMPLICATIONS AND PROGNOSIS

IC/BPS is not unlike other chronic pain conditions in that, although a clear link has not been established with higher mortality, this condition is certainly associated with significant disability, decreased quality of life, and significant mental health morbidity. The economic impact of the disability associated with IC/BPS is similar to that of fibromyalgia, low back pain, rheumatoid arthritis, and peripheral neuropathy. Suicidal ideation is a reality in this patient population, with a reported prevalence as high as 11–23%.

For most patients, IC/BPS onset is subacute, with continuous development of the classic symptom complex over a short time and rapid (within 5 years) progression to its final stage. Symptoms then continue to wax and wane without significant overall change in symptomatology for the majority of patients. However, spontaneous improvement and/or resolution occurs in some patients, while a small subset experience subsequent deterioration to a small-capacity, fibrotic, noncompliant bladder (“end-stage bladder”). A multimodal approach to therapy, interdisciplinary involvement in patient care, particular attention to psychosocial parameters, and check-ins on mental health are important aspects of ongoing care.

GLOBAL CONSIDERATIONS

Significant challenges have been encountered in confirming the prevalence of IC/BPS, particularly globally. Prevalence estimates have ranged widely from as low as 3.5 per 100,000 women in a study of a Japanese population to as high as 20,000 per 100,000 in a self-report questionnaire study of a U.S. population. Despite these challenges, it has been recognized that IC/BPS is not simply a disease of the global West. Although robust epidemiologic studies outside North America, Europe, and some regions of Asia are lacking, it is presumed that this disease occurs at similar rates globally. This presumption may be extrapolated from epidemiologic studies of a related population and its male counterpart: CP/CPPS. These studies have shown rates of CP/CPPS in African and Asian populations that are similar to rates in North American populations.

There is no evidence to suggest that IC/BPS is phenotypically distinct in various geographic regions. Thus, this condition should be diagnosed and treated in the same ways globally. Given that its diagnosis of exclusion is based largely on history and physical examination and its treatment is based on a minimally invasive algorithm, with the focus on the patient's clinical phenotype and the initial implementation of conservative therapeutic measures, IC/BPS can be well managed even in resource-poor settings. As with many poorly understood and difficult-to-treat conditions, the greatest barrier to its diagnosis and treatment may perhaps be its recognition.

FURTHER READING

- CLEMENS JQ et al: Urologic chronic pelvic pain syndrome: Insights from the MAPP Research Network. *Nat Rev Urol* 16:187, 2019.
- COX A et al: CUA guideline: Diagnosis and treatment of interstitial cystitis/bladder pain syndrome. *Can Urol Assoc J* 10:E136, 2016.
- HANNO PM et al: AUA guideline for the diagnosis and treatment of interstitial cystitis/bladder pain syndrome. *J Urol* 185:2162, 2011.
- HANNO P et al: Incontinence, in *International Consultation on Incontinence, September 2016*, vol 2, 6th ed, P Abrams et al (eds). Tokyo, ICUD ICS, 2017, pp 2203–2301.
- VAN DE MERWE JP et al: Diagnostic criteria, classification, and nomenclature for painful bladder syndrome/interstitial cystitis: An ESSIC proposal. *Eur Urol* 53:60, 2008.

52

Azotemia and Urinary Abnormalities

David B. Mount



Normal kidney functions occur through numerous cellular processes to maintain body homeostasis. Disturbances in any of these functions can lead to abnormalities that may be detrimental to survival. Clinical manifestations of these disorders depend on the pathophysiology of renal injury and often are identified as a complex of symptoms, abnormal physical findings, and laboratory changes that constitute specific syndromes. These renal syndromes (Table 52-1) may arise from systemic illness or as primary renal disease. Nephrologic syndromes usually consist of several elements that reflect the underlying pathologic processes, typically including one or more of the following: (1) reduction in glomerular filtration rate (GFR), (2) abnormalities of urine sediment (red blood cells [RBCs], white blood cells [WBCs], casts, and crystals), (3) abnormal urinary excretion of serum proteins (proteinuria), (4) disturbances in urine volume (oliguria, anuria, polyuria), (5) presence of hypertension and/or expanded total body fluid volume (edema), (6) electrolyte abnormalities, and (7) in some syndromes, fever/pain. The specific combination of these findings should permit identification of one of the major nephrologic syndromes (Table 52-1) and allow differential diagnoses to be narrowed so that the appropriate diagnostic and therapeutic course can be determined. All these syndromes and their associated diseases are discussed in more detail in subsequent chapters. This chapter focuses on several aspects of renal abnormalities that are critically important for distinguishing among those processes: (1) reduction in GFR, (2) alterations of the urinary sediment and/or protein excretion, and (3) abnormalities of urinary volume.

AZOTEMIA

ASSESSMENT OF GFR

Monitoring the GFR is important in both hospital and outpatient settings, and several different methodologies are available. GFR is the primary metric for kidney “function,” and its direct measurement involves administration of a radioactive isotope (such as inulin or iothalamate) that is filtered at the glomerulus into the urinary space but is neither reabsorbed nor secreted throughout the tubule. GFR—i.e., the clearance of inulin or iothalamate in milliliters per minute—is calculated from the rate of appearance of the isotope in the urine over several hours. In most clinical circumstances, direct GFR measurement is not feasible, and the plasma creatinine level is used as a surrogate to estimate GFR. Plasma creatinine (P_{Cr}) is the most widely used marker for GFR, which is related directly to urine creatinine (U_{Cr}) excretion and inversely to P_{Cr} . On the basis of this relationship (with some important caveats, as discussed below), GFR will fall in roughly inverse proportion to the rise in P_{Cr} . Failure to account for GFR reductions in drug dosing can lead to significant morbidity and death from drug toxicities (e.g., digoxin, imipenem). In the outpatient setting, P_{Cr} serves as an estimate for GFR (although much less accurate; see below). In patients with chronic progressive renal disease, there is an approximately linear relationship between $1/P_{Cr}$ (y axis) and time (x axis). The slope of that line will remain constant for an individual; when values deviate, an investigation for a superimposed acute process (e.g., volume depletion, drug reaction) should be initiated. Signs and symptoms of uremia, the clinical symptom complex associated with renal failure, develop at significantly different levels of P_{Cr} depending on the patient (size, age, and sex), underlying renal disease, existence of concurrent diseases, and true GFR. Generally, patients do not develop symptomatic uremia until renal insufficiency is severe (GFR <15 mL/min).

A significantly reduced GFR (either acute or chronic) is usually reflected in a rise in P_{Cr} leading to retention of nitrogenous waste products (defined as azotemia) such as urea. Azotemia may result from

TABLE 52-1 Initial Clinical and Laboratory Database for Defining Major Syndromes in Nephrology

SYNDROME	IMPORTANT CLUES TO DIAGNOSIS	COMMON FINDINGS	CHAP(S). DISCUSSING DISEASE-CAUSING SYNDROME
Acute or rapidly progressive renal failure	Anuria	Hypertension, hematuria	310, 314, 316, 319
	Oliguria	Proteinuria, pyuria	
	Documented recent decline in GFR	Casts, edema	
Acute nephritis	Hematuria, RBC casts	Proteinuria	314
	Azotemia, reduced GFR, oliguria	Pyuria	
	Edema, hypertension	Circulatory congestion	
Chronic renal failure	Azotemia for >3 months	Proteinuria, casts	311
	Symptoms or signs of uremia, (late manifestation), casts	Hypocalcemia, hyperphosphatemia, hyperparathyroidism	
	Symptoms or signs of renal osteodystrophy	Polyuria, nocturia	
	Kidneys reduced in size bilaterally	Edema, hypertension	
	Broad casts in urinary sediment	Hyperkalemia, metabolic acidosis	
Nephrotic syndrome	Proteinuria, with >3.5 g/24 h per 1.73 m ²	Casts	314
	Hypoalbuminemia	Lipiduria	
	Edema	Hypercoagulable state	
	Hyperlipidemia		
Asymptomatic urinary abnormalities	Hematuria		314
	Proteinuria (below nephrotic range)		
	Sterile pyuria, casts		
Urinary tract infection/pyelonephritis	Bacteriuria, with >10 ⁵ cfu/mL	Hematuria	135
	Other infectious agent documented in urine	Mild azotemia and reduced GFR	
	Pyuria, leukocyte casts	Mild proteinuria	
	Frequency, urgency	Fever	
	Bladder tenderness, flank tenderness		
Renal tubular defects	Electrolyte disorders	Hematuria	315, 316
	Polyuria, nocturia	"Tubular" proteinuria (<1 g/24 h)	
	Renal calcification	Enuresis	
	Large kidneys	Electrolyte and/or acid-base abnormalities	
	Renal transport defects	Other electrolyte issues, e.g., hypomagnesemia	
Hypertension	Systolic/diastolic hypertension	Proteinuria	277, 317
		Casts	
		Azotemia	
Nephrolithiasis	Previous history of stone passage or removal	Hematuria	318
	Previous history of stone seen by x-ray	Pyuria	
	Renal colic	Frequency, urgency	
Urinary tract obstruction	Azotemia, oliguria, anuria	Hematuria	319
	Polyuria, nocturia, urinary retention	Pyuria	
	Slowing of urinary stream	Enuresis, dysuria	
	Large prostate, large kidneys		
	Flank tenderness, full bladder after voiding		

Abbreviations: cfu, colony-forming units; GFR, glomerular filtration rate; RBC, red blood cell.

reduced renal perfusion, intrinsic renal disease, or postrenal processes (ureteral obstruction; see below and Fig. 52-1). Precise determination of GFR is problematic, as both commonly measured indices (urea and creatinine) have characteristics that affect their accuracy as markers of clearance. Urea clearance may underestimate GFR significantly because of urea reabsorption by the tubule. In contrast, creatinine is derived from muscle metabolism of creatine, and its generation varies little from day to day.

Creatinine clearance (CrCl), an approximation of GFR, is measured from plasma and urinary creatinine excretion rates for a defined period (usually 24 h) and is expressed in milliliters per minute: CrCl = $(U_{\text{vol}} \times U_{\text{Cr}})/(P_{\text{Cr}} \times T_{\text{min}})$. The "adequacy" or "completeness" of the urinary collection is estimated by the urinary volume and creatinine content; creatinine is produced from muscle and excreted at a relatively constant

rate. For a 20- to 50-year-old man, creatinine excretion should be 18.5–25.0 mg/kg body weight; for a woman of the same age, it should be 16.5–22.4 mg/kg body weight. For example, an 80-kg man should excrete between ~1500 and 2000 mg of creatinine in an "adequate" collection. Creatinine is useful for estimating GFR because it is a small, freely filtered solute that is not reabsorbed by the tubules. P_{Cr} levels can increase acutely from dietary ingestion of cooked meat, however, and creatinine can be secreted into the proximal tubule through an organic cation pathway (especially in advanced progressive chronic kidney disease [CKD]), leading to overestimation of GFR. When a timed collection for CrCl is not available, decisions about drug dosing must be based on P_{Cr} alone. Two formulas are used widely to estimate kidney function from P_{Cr} : (1) Cockcroft-Gault and (2) four-variable MDRD (Modification of Diet in Renal Disease).

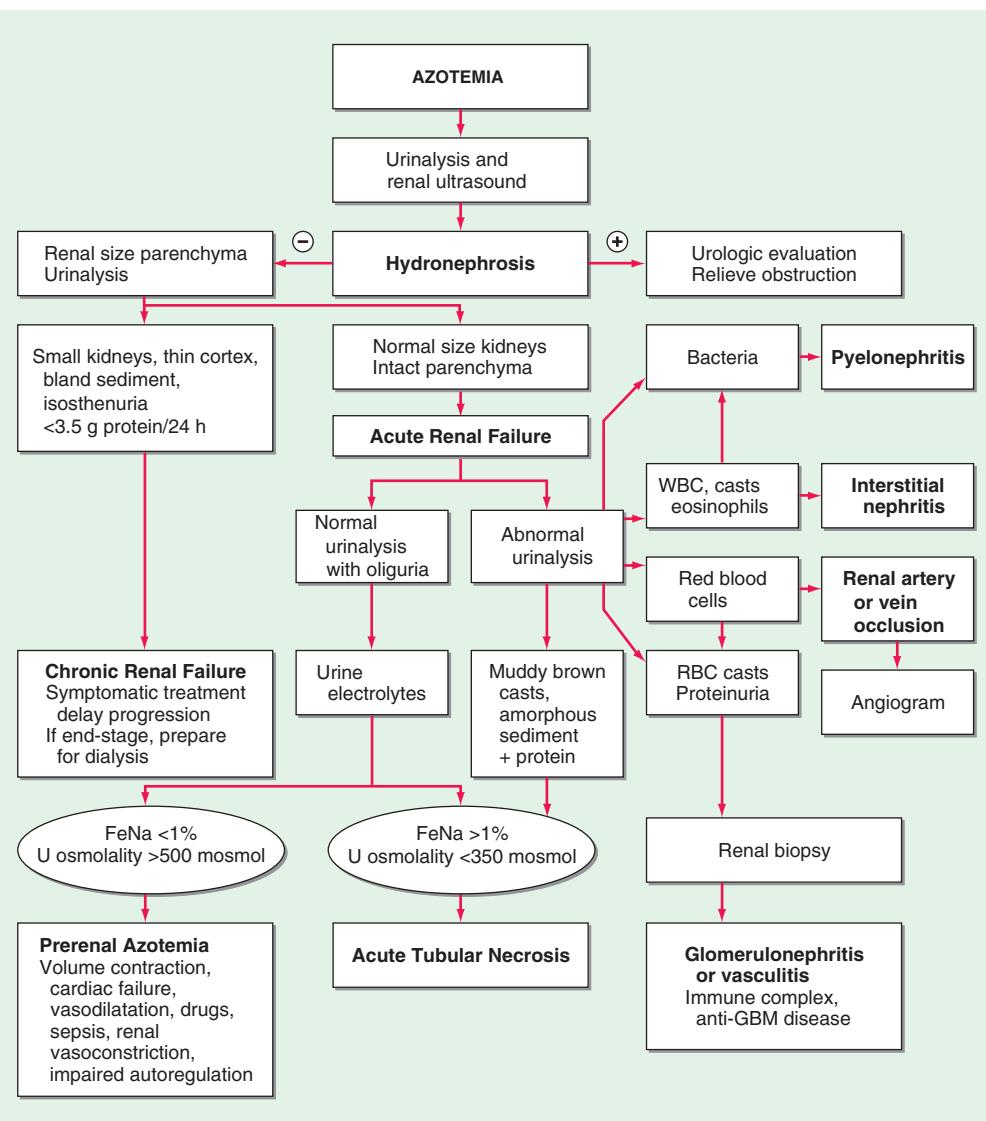


FIGURE 52-1 Approach to the patient with azotemia. FeNa, fractional excretion of sodium; GBM, glomerular basement membrane; RBC, red blood cell; U, urine; WBC, white blood cell.

Cockcroft-Gault:

$$\text{CrCl}(\text{mL/min}) = \frac{(140 - \text{age}) \times \text{Lean Body Weight (kg)}}{\text{Serum Creatinine (mg/dL)} \times 72} \times 0.85 \text{ if female}$$

$$\text{MDRD: eGFR (mL/min per } 1.73 \text{ m}^2\text{)} = 186.3 \times P_{\text{Cr}}^{e^{-1.154}} \times \text{age}^{e^{-0.203}} \times (0.742 \text{ if female}) \times (1.21 \text{ if black}).$$

Numerous websites are available to assist with these calculations (www.kidney.org/professionals/kdoqi/gfr_calculator.cfm). A newer Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) estimated GFR (eGFR), which was developed by pooling several cohorts with and without kidney disease who had data on directly measured GFR, appears to be more accurate:

$$\text{CKD-EPI: eGFR} = 141 \times \min(P_{\text{Cr}}/k, 1)^a \times \max(P_{\text{Cr}}/k, 1)^{-1.209} \times 0.993^{\text{Age}} \times 1.018 \text{ [if female]} \times 1.159 \text{ [if black]},$$

where P_{Cr} is plasma creatinine, k is 0.7 for females and 0.9 for males, a is -0.329 for females and -0.411 for males, \min indicates the minimum of P_{Cr}/k or 1, and \max indicates the maximum of P_{Cr}/k or 1 (<https://www.mdcalc.com/ckd-epi-equations-glomerular-filtration-rate-gfr>).

There are limitations to all creatinine-based estimates of GFR. Each equation, along with 24-h urine collection for measurement of creatinine clearance, is based on the assumption that the patient is in *steady state*, without daily increases or decreases in P_{Cr} as a result of rapidly changing GFR. The MDRD equation is better correlated with true GFR when the GFR is <60 mL/min per 1.73 m^2 . The gradual loss of muscle from chronic illness, chronic use of glucocorticoids, or malnutrition can mask significant changes in GFR with small or imperceptible changes in P_{Cr} .

The coefficient of 1.159 in the CKD-EPI equation to adjust for self-reported black race reflects that measured GFR was 16% higher in blacks than nonblacks with similar age, sex, and creatinine in the data set used to develop the equation. Race is a social rather than a biological construct, for which reason the use of the “race modifier” in calculating eGFR using CKD-EPI and other equations has come under scrutiny. In particular, given the implications of utilizing self-reported race to modify clinical laboratory results, many medical centers have recently stopped reporting eGFRs that have been calculated using a race modifier. This change is projected to have positive consequences, in particular, improved access to waitlisting for renal transplantation in black patients at an earlier stage of CKD. Potential negative consequences include “overdiagnosis” of CKD, inadequate or inaccurate dosing of drugs that are eliminated through the kidney (e.g.,

metformin), reduced access to imaging modalities for black patients with CKD with a lower reported eGFR, and reductions in living kidney donation among blacks. These and the other limitations in creatinine-based eGFR have led to the development of alternative methods for estimating GFR.

Cystatin C, a member of the cystatin superfamily of cysteine protease inhibitors, is produced at a relatively constant rate from all nucleated cells. Serum cystatin C has been proposed to be a more sensitive marker of early GFR decline than is P_{Cr} , with lesser effects of muscle mass on circulating levels; however, cystatin C levels are influenced by the patient's sex and the presence of diabetes mellitus, smoking, and inflammation. To the extent that cystatin C-based calculation of eGFR is less affected by self-reported race and muscle mass, it is an increasingly important adjunct to creatinine-based eGFR.

APPROACH TO THE PATIENT

Azotemia

Once GFR reduction has been established, the physician must decide if it represents acute or chronic renal injury. The clinical circumstances, history, and laboratory data often make this an easy distinction. However, the laboratory abnormalities characteristic of chronic renal failure, including anemia, hypocalcemia, and hyperphosphatemia, are also often present in patients presenting with acute renal failure. Radiographic evidence of renal osteodystrophy ([Chap. 311](#)) can be seen only in chronic renal failure but is a very late finding, typically in patients with end-stage renal disease (ESRD) maintained on dialysis. The urinalysis and renal ultrasound can facilitate distinguishing acute from chronic renal failure. An approach to the evaluation of azotemic patients is shown in Fig. 52-1. Patients with advanced chronic renal insufficiency often have some proteinuria, nonconcentrated urine (isosthenuria; isosmotic with plasma), and small kidneys on ultrasound, characterized by increased echogenicity and cortical thinning. Treatment should be directed toward slowing the progression of renal disease and providing symptomatic relief for edema, acidosis, anemia, and hyperphosphatemia, as discussed in [Chap. 311](#). Acute renal failure ([Chap. 310](#)) can result from processes that affect blood flow and glomerular perfusion (prerenal azotemia), intrinsic renal diseases (affecting small vessels, glomeruli, or tubules), or postrenal processes (obstruction of urine flow in ureters, bladder, or urethra) ([Chap. 319](#)).

PRERENAL FAILURE

Decreased renal perfusion accounts for 40–80% of cases of acute renal failure and, if appropriately treated, is readily reversible. The etiologies of prerenal azotemia include any cause of decreased circulating blood volume (gastrointestinal hemorrhage, burns, diarrhea, diuretics), volume sequestration (pancreatitis, peritonitis, rhabdomyolysis), or decreased effective arterial volume (cardiogenic shock, sepsis). Renal and glomerular perfusion also can be affected by reductions in cardiac output from peripheral vasodilation (sepsis, drugs) or profound renal vasoconstriction (severe heart failure, hepatorenal syndrome, agents such as nonsteroidal anti-inflammatory drugs [NSAIDs]). True or “effective” arterial hypovolemia leads to a fall in mean arterial pressure, which in turn triggers a series of neural and humoral responses, including activation of the sympathetic nervous and renin-angiotensin-aldosterone systems and vasopressin (AVP) release. GFR is maintained by prostaglandin-mediated dilatation of afferent arterioles and angiotensin II-mediated constriction of efferent arterioles. Once the mean arterial pressure falls below 80 mmHg, GFR declines steeply.

Blockade of prostaglandin production by NSAIDs can result in severe vasoconstriction and acute renal failure. Blocking angiotensin action with angiotensin-converting enzyme (ACE) inhibitors

TABLE 52-2 Laboratory Findings in Acute Renal Failure

INDEX	PRERENAL AZOTEMIA	OLIGURIC ACUTE RENAL FAILURE
BUN/ P_{Cr} ratio	>20:1	10–15:1
Urine sodium U_{Na} , meq/L	<20	>40
Urine osmolality, mosmol/L H_2O	>500	<350
Fractional excretion of sodium ^a	<1%	>2%
Urine/plasma creatinine U_{Cr}/P_{Cr}	>40	<20
Urinalysis (casts)	None or hyaline/granular	Muddy brown

$$^aFE_{Na} = \frac{U_{Na} \times P_{Cr} \times 100}{P_{Na} \times U_{Cr}}$$

Abbreviations: BUN, blood urea nitrogen; P_{Cr} , plasma creatinine concentration; P_{Na} , plasma sodium concentration; U_{Cr} , urine creatinine concentration; U_{Na} , urine sodium concentration.

or angiotensin receptor blockers (ARBs) decreases efferent arteriolar tone and in turn decreases glomerular capillary perfusion pressure. Patients taking NSAIDs and/or ACE inhibitors/ARBs are most susceptible to hemodynamically mediated acute renal failure when blood volume or arterial perfusion pressure is reduced for any reason; under these circumstances, preservation of GFR is dependent on afferent vasodilation due to prostaglandins and efferent vasoconstriction due to angiotensin II. Patients with bilateral renal artery stenosis (or stenosis in a solitary kidney) can also be dependent on efferent arteriolar vasoconstriction for maintenance of glomerular filtration pressure and are particularly susceptible to a precipitous decline in GFR when given ACE inhibitors or ARBs.

Prolonged renal hypoperfusion may lead to acute tubular necrosis (ATN), an intrinsic renal disease that is discussed below. The urinalysis and urinary electrolyte measurements can be useful in distinguishing prerenal azotemia from ATN ([Table 52-2](#)). The urine Na and osmolality of patients with prerenal azotemia can be predicted from the stimulatory actions of norepinephrine, angiotensin II, AVP, aldosterone, and low tubule fluid flow rate. In prerenal conditions, the tubules are intact, leading to a concentrated urine (>500 mosmol), avid Na retention (urine Na concentration, <20 mmol/L; fractional excretion of Na [FE_{Na}], <1%), and $U_{Cr}/P_{Cr} >40$ ([Table 52-2](#)). The FE_{Na} is typically >1% in ATN, but may be <1% in patients with milder, nonoliguric ATN (e.g., from rhabdomyolysis) and in patients with underlying “prerenal” disorders, such as congestive heart failure (CHF) or cirrhosis or hepatorenal syndrome. The prerenal urine sediment is usually normal or has hyaline and granular casts, whereas the sediment of ATN usually is filled with cellular debris, tubular epithelial casts, and dark (muddy brown) granular casts. The measurement of urinary biomarkers associated with tubular injury is a promising technique to detect subclinical ATN and/or help further diagnose the exact cause of acute renal failure.

POSTRENAL AZOTEMIA

Urinary tract obstruction accounts for <5% of cases of acute renal failure but is usually reversible and must be ruled out early in the evaluation ([Fig. 52-1](#)). Since a single kidney is capable of adequate clearance, complete obstructive acute renal failure requires obstruction at the urethra or bladder outlet, bilateral ureteral obstruction, or unilateral obstruction in a patient with a single functioning kidney. Obstruction is usually diagnosed by the presence of ureteral and renal pelvic dilation on renal ultrasound. However, early in the course of obstruction or if the ureters are unable to dilate (e.g., encasement by pelvic or periureteral tumors or by retroperitoneal fibrosis), the ultrasound examination may be negative. Other

imaging, such as a furosemide renogram (MAG3 nuclear medicine study), may be required to better define the presence or absence of obstructive uropathy. The specific urologic conditions that cause obstruction are discussed in [Chap. 319](#).

INTRINSIC RENAL DISEASE

When prerenal and postrenal azotemia have been excluded as etiologies of renal failure, an intrinsic parenchymal renal disease is present. Intrinsic renal disease can arise from processes involving large renal vessels, intrarenal microvasculature and glomeruli, or the tubulointerstitium. Ischemic and toxic ATN account for ~90% of cases of acute intrinsic renal failure. As outlined in Fig. 52-1, the clinical setting and urinalysis are helpful in separating the possible etiologies. Prerenal azotemia and ATN are part of a spectrum of renal hypoperfusion; evidence of structural tubule injury is present in ATN, whereas prompt reversibility occurs with prerenal azotemia upon restoration of adequate renal perfusion. Thus, ATN often can be distinguished from prerenal azotemia by urinalysis and urine electrolyte composition (Table 52-2 and Fig. 52-1). Ischemic ATN is observed most frequently in patients who have undergone major surgery, trauma, severe hypovolemia, overwhelming sepsis, or extensive burns. Nephrotoxic ATN complicates the administration of many common medications, usually by inducing a combination of intrarenal vasoconstriction, direct tubule toxicity, and/or tubular obstruction. The kidney is vulnerable to toxic injury by virtue of its rich blood supply (25% of cardiac output) and its ability to concentrate and metabolize toxins. A diligent search for hypotension and nephrotoxins usually uncovers the specific etiology of ATN. Discontinuation of nephrotoxins and stabilization of blood pressure often suffice without the need for dialysis, with ongoing regeneration of tubular cells. [An extensive list of potential drugs and toxins implicated in ATN is found in Chap. 310.](#)

Processes involving the tubules and interstitium can lead to acute kidney injury (AKI), a subtype of acute renal failure. These processes include drug-induced interstitial nephritis (especially by antibiotics, NSAIDs, and diuretics), severe infections (both bacterial and viral), systemic diseases (e.g., systemic lupus erythematosus), and systemic disorders (e.g., sarcoidosis, Sjögren's syndrome, lymphoma, or leukemia). A list of drugs associated with allergic interstitial nephritis is found in [Chap. 316](#). Urinalysis usually shows mild to moderate proteinuria, hematuria, and pyuria (~75% of cases) and occasionally WBC casts. The finding of RBC casts in interstitial nephritis has been reported but should prompt a search for glomerular diseases (Fig. 52-1). Occasionally, renal biopsy will be needed to distinguish among these possibilities. The classic sediment finding in allergic interstitial nephritis is a predominance (>10%) of urinary eosinophils with Wright's or Hansel's stain; however, urinary eosinophils can be increased in several other causes of AKI, such that measurement of urine eosinophils has no diagnostic utility in renal disease.

Occlusion of large renal vessels, including arteries and veins, is an uncommon cause of acute renal failure. A significant reduction in GFR by this mechanism suggests bilateral processes or, in a patient with a single functioning kidney, a unilateral process. In patients with preexisting renal artery stenosis, a substantial renal collateral circulation can develop over time and sustain renal perfusion—typically not enough to sustain glomerular filtration—in the event of total renal artery occlusion. Renal arteries can be occluded with atheroemboli, thromboemboli, in situ thrombosis, aortic dissection, or vasculitis. Atheroembolic renal failure can occur spontaneously but most often is associated with recent aortic instrumentation. The emboli are cholesterol-rich and lodge in medium and small renal arteries, with a consequent eosinophil-rich inflammatory reaction. Patients with atheroembolic acute renal failure often have a normal urinalysis, but the urine may contain eosinophils and casts. The diagnosis can be confirmed by renal biopsy, but this procedure is often unnecessary when other stigmata

of atheroemboli are present (livedo reticularis, distal peripheral infarcts, eosinophilia). Renal artery thrombosis may lead to mild proteinuria and hematuria, whereas renal vein thrombosis typically occurs in the context of heavy proteinuria and hematuria. These vascular complications often require angiography for confirmation and are discussed in [Chap. 317](#).

Diseases of the glomeruli (glomerulonephritis and vasculitis) and the renal microvasculature (hemolytic-uremic syndromes, thrombotic thrombocytopenic purpura, and malignant hypertension) usually present with various combinations of glomerular injury: proteinuria, hematuria, reduced GFR, and alterations of sodium excretion that lead to hypertension, edema, and circulatory congestion (acute nephritic syndrome). These findings may occur as primary renal diseases or as renal manifestations of systemic diseases. The clinical setting and other laboratory data help distinguish primary renal diseases from systemic diseases. The finding of RBC casts in the urine is an indication for early renal biopsy (Fig. 52-1), as the pathologic pattern has important implications for diagnosis, prognosis, and treatment. Hematuria without RBC casts can also be an indication of glomerular disease, since RBC casts are highly specific but very insensitive for glomerulonephritis. The specificity of urine microscopy can be enhanced by examining urine with a phase contrast microscope capable of detecting dysmorphic red cells ("acanthocytes") that are associated with glomerular disease. This evaluation is summarized in [Fig. 52-2](#). A detailed discussion of glomerulonephritis and diseases of the microvasculature is found in [Chap. 316](#).

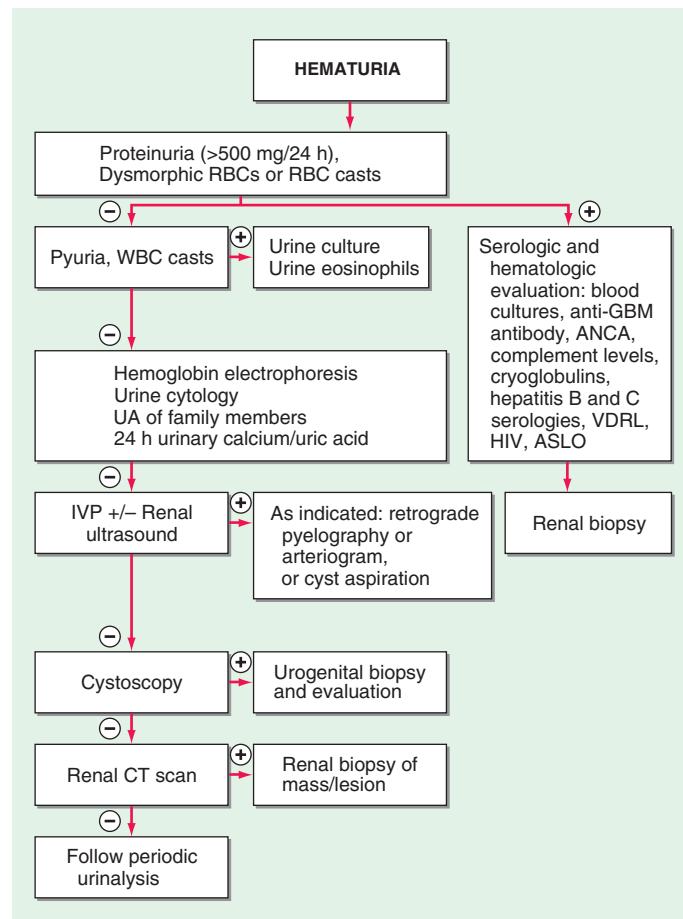


FIGURE 52-2 Approach to the patient with hematuria. ANCA, antineutrophil cytoplasmic antibody; ASLO, antistreptolysin O; CT, computed tomography; GBM, glomerular basement membrane; IVP, intravenous pyelography; RBC, red blood cell; UA, urinalysis; VDRL, Venereal Disease Research Laboratory; WBC, white blood cell.

OLIGURIA AND ANURIA

Oliguria refers to a 24-h urine output <400 mL, and *anuria* is the complete absence of urine formation (<100 mL). Anuria can be caused by complete bilateral urinary tract obstruction; a vascular catastrophe (dissection or arterial occlusion); renal vein thrombosis; acute cast nephropathy in myeloma; renal cortical necrosis; severe ATN; combined therapy with NSAIDs, ACE inhibitors, and/or ARBs; and hypovolemic, cardiogenic, or septic shock. Oliguria is never normal, since at least 400 mL of maximally concentrated urine must be produced to excrete the obligate daily osmolar load. *Nonoliguria* refers to urine output >400 mL/d in patients with acute or chronic azotemia. With nonoliguric ATN, disturbances of potassium and hydrogen balance are less severe than in oliguric patients, and recovery to normal renal function is usually more rapid.

ABNORMALITIES OF THE URINE

■ PROTEINURIA

The evaluation of proteinuria is shown schematically in Fig. 52-3 and typically is initiated after detection of proteinuria by dipstick examination. The dipstick measurement detects only albumin and gives false-positive results at pH >7.0 or when the urine is very concentrated or contaminated with blood. Because the dipstick relies on urinary albumin concentration, a very dilute urine may obscure significant proteinuria on dipstick examination. Quantification of urinary albumin on a spot urine sample (ideally from a first morning void) by measurement of an albumin-to-creatinine ratio (ACR) is helpful in approximating a 24-h albumin excretion rate (AER), where ACR (mg/g) \approx AER (mg/24 h). Furthermore, proteinuria that is not predominantly due to albumin will be missed by dipstick screening. This information is particularly important for the detection of Bence-Jones proteins in the urine of patients with multiple myeloma. Tests to measure total urine protein concentration accurately rely on precipitation

with sulfosalicylic or trichloroacetic acid (Fig. 52-3). As with albuminuria, the ratio of protein to creatinine in a random, "spot" urine can also provide a rough estimate of protein excretion; for example, a protein/creatinine ratio of 3.0 correlates to ~3.0 g of proteinuria per day. Formal assessment of urinary protein excretion requires a 24-h urine protein collection (see "Measurement of GFR," above).

The magnitude of proteinuria and its composition in the urine depend on the mechanism of renal injury that leads to protein losses. Both charge and size selectivity normally prevent virtually all plasma albumin, globulins, and other high-molecular-weight proteins from crossing the glomerular wall; however, if this barrier is disrupted, plasma proteins may leak into the urine (glomerular proteinuria; Fig. 52-3). Smaller proteins (<20 kDa) are freely filtered but are readily reabsorbed by the proximal tubule. Typically, healthy individuals excrete <150 mg/d of total protein and <30 mg/d of albumin. However, even at albuminuria levels <30 mg/d, risk for progression to overt nephropathy or subsequent cardiovascular disease is increased. The remainder of the protein in the urine is secreted by the tubules (Tamm-Horsfall, IgA, and urokinase) or represents small amounts of filtered β_2 -microglobulin, apoproteins, enzymes, and peptide hormones. Another mechanism of proteinuria entails excessive production of an abnormal protein that exceeds the capacity of the tubule for reabsorption. This situation most commonly occurs with plasma cell dyscrasias, such as multiple myeloma, amyloidosis, and lymphomas, that are associated with monoclonal production of immunoglobulin light chains.

The normal glomerular endothelial cell forms a barrier composed of pores of ~100 nm that retain blood cells but offer little impediment to passage of most proteins. The glomerular basement membrane traps most large proteins (>100 kDa), and the foot processes of epithelial cells (podocytes) cover the urinary side of the glomerular basement membrane and produce a series of narrow channels (slit diaphragms) to allow molecular passage of small solutes and water but not proteins. Some glomerular diseases, such as minimal change disease, cause fusion of glomerular epithelial cell foot processes, resulting in predominantly "selective" (Fig. 52-3) loss of albumin. Other glomerular diseases can present with disruption of the basement membrane and slit diaphragms (e.g., by immune complex deposition), resulting in losses of albumin and other plasma proteins. The fusion of foot processes causes increased pressure across the capillary basement membrane, resulting in areas with larger pore sizes (and more severe "nonselective" proteinuria) (Fig. 52-3).

When the total daily urinary excretion of protein is >3.5 g, hypoalbuminemia, hyperlipidemia, and edema (nephrotic syndrome; Fig. 52-3) are often present as well. However, total daily urinary protein excretion >3.5 g can occur without the other features of the nephrotic syndrome in a variety of other renal diseases, including diabetes (Fig. 52-3). Plasma cell dyscrasias (multiple myeloma) can be associated with large amounts of excreted light chains in the urine, which may not be detected by dipstick. The light chains are filtered by the glomerulus and overwhelm the reabsorptive capacity of the proximal tubule. Renal failure from these disorders occurs through a variety of mechanisms, including but not limited to proximal tubule injury, tubule obstruction (cast nephropathy), amyloid deposition, and light chain deposition (Chap. 316). The specific renal lesion is dictated by the sequence and structural

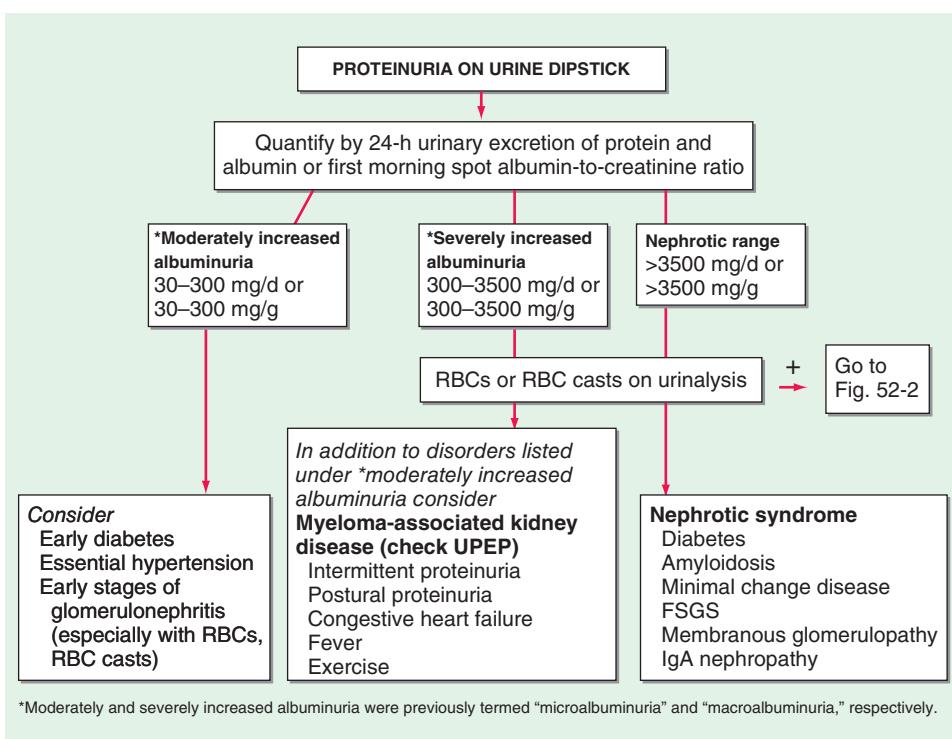


FIGURE 52-3 Approach to the patient with proteinuria. Investigation of proteinuria is often initiated by a positive dipstick on routine urinalysis. Conventional dipsticks detect predominantly albumin and provide a semiquantitative assessment (trace, 1+, 2+, or 3+), which is influenced by urinary concentration as reflected by urine specific gravity (minimum, <1.005; maximum, 1.030). However, more exact determination of proteinuria should employ a spot morning protein/creatinine ratio (mg/g) or a 24-h urine collection (mg/24 h). FSGS, focal segmental glomerulosclerosis; RBC, red blood cell; UPEP, urine protein electrophoresis.

characteristics of the monoclonal light chain; however, not all excreted light chains are nephrotoxic.

Hypoalbuminemia in nephrotic syndrome occurs through excessive urinary losses and increased proximal tubule catabolism of filtered albumin. Edema results from renal sodium retention and reduced plasma oncotic pressure, which favors fluid movement from capillaries to interstitium. To compensate for the perceived decrease in effective intravascular volume, activation of the renin-angiotensin system, stimulation of AVP, and activation of the sympathetic nervous system take place, promoting continued renal salt and water reabsorption and progressive edema. Filtered proteases, normally retained by the glomerular filtration barrier, can also directly activate sodium reabsorption by the epithelial Na channels in principal cells (ENaC) in nephrotic syndrome. Despite these changes, hypertension is uncommon in primary kidney diseases resulting in the nephrotic syndrome (Fig. 52-3 and [Chap. 314](#)). The urinary loss of regulatory proteins and changes in hepatic synthesis contribute to the other manifestations of the nephrotic syndrome. A hypercoagulable state may arise from urinary losses of antithrombin III, reduced serum levels of proteins S and C, hyperfibrinogenemia, and enhanced platelet aggregation. Hypercholesterolemia may be severe and results from increased hepatic lipoprotein synthesis. Loss of immunoglobulins contributes to an increased risk of infection. Many diseases (some listed in Fig. 52-3) and drugs can cause the nephrotic syndrome; a complete list is found in [Chap. 314](#).

■ HEMATURIA, PYURIA, AND CASTS

Isolated hematuria without proteinuria, other cells, or casts is often indicative of bleeding from the urinary tract. Hematuria is defined as two to five RBCs per high-power field (HPF) and can be detected by dipstick. A false-positive dipstick for hematuria (where no RBCs are seen on urine microscopy) may occur when myoglobinuria is present, often in the setting of rhabdomyolysis. Common causes of isolated hematuria include stones, neoplasms, tuberculosis, trauma, and prostatitis. Gross hematuria with blood clots usually is not an intrinsic renal process; rather, it suggests a postrenal source in the urinary collecting system. Evaluation of patients presenting with microscopic hematuria is outlined in Fig. 52-2. A single urinalysis with hematuria is common and can result from menstruation, viral illness, allergy, exercise, or mild trauma. Persistent or significant hematuria (>3 RBCs/HPF on three urinalyses, a single urinalysis with >100 RBCs, or gross hematuria) is associated with significant renal or urologic lesions in 9.1% of cases. The level of suspicion for urogenital neoplasms in patients with isolated painless hematuria and nondysmorphic RBCs increases with age. Neoplasms are rare in the pediatric population, and isolated hematuria is more likely to be “idiopathic” or associated with a congenital anomaly. Hematuria with pyuria and bacteriuria is typical of infection and should be treated with antibiotics after appropriate cultures. Acute cystitis or urethritis in women can cause gross hematuria. Hypercalcuria and hyperuricosuria are also risk factors for unexplained isolated hematuria in both children and adults. In some of these patients (50–60%), reducing calcium and uric acid excretion through dietary interventions can eliminate the microscopic hematuria.

Isolated microscopic hematuria can be a manifestation of glomerular diseases. The RBCs of glomerular origin are often dysmorphic when examined by phase-contrast microscopy. Irregular shapes of RBCs may also result from pH and osmolarity changes produced along the distal nephron. Observer variability in detecting dysmorphic RBCs is common. The most common etiologies of isolated glomerular hematuria are IgA nephropathy, hereditary nephritis, and thin basement membrane disease. IgA nephropathy and hereditary nephritis can lead to episodic gross hematuria. A family history of renal failure is often present in hereditary nephritis, and patients with thin basement membrane disease often have family members with microscopic hematuria. A renal biopsy is needed for the definitive diagnosis of these disorders, which are discussed in more detail in [Chap. 314](#). Hematuria with dysmorphic RBCs, RBC casts, and protein excretion >500 mg/d is virtually diagnostic of glomerulonephritis. RBC casts form as RBCs that enter the tubule fluid and become trapped in a cylindrical mold of gelled Tamm-Horsfall protein. Even in the absence of azotemia,

these patients should undergo serologic evaluation and renal biopsy as outlined in Fig. 52-2.

Isolated pyuria is unusual since inflammatory reactions in the kidney or collecting system also are associated with hematuria. The presence of bacteria suggests infection, and WBC casts with bacteria are indicative of pyelonephritis; “sterile pyuria” with negative urinary bacterial cultures can be seen in urogenital tuberculosis. WBCs and/or WBC casts also may be seen in acute glomerulonephritis as well as in tubulointerstitial processes such as interstitial nephritis and transplant rejection.

Casts can be seen in chronic renal diseases. Degenerated cellular casts called *waxy casts* or *broad casts* (arising in the dilated tubules that have undergone compensatory hypertrophy in response to reduced renal mass) may be seen in the urine.

ABNORMALITIES OF URINE VOLUME

■ POLYURIA

By history, it is often difficult for patients to distinguish urinary frequency (often of small volumes) from true polyuria (>3 L/d), and a quantification of volume by 24-h urine collection may be needed ([Fig. 52-4](#)). Polyuria results from two potential mechanisms: (1) excretion of nonabsorbable solutes (such as glucose) or (2) excretion of water (usually from a defect in AVP production or renal responsiveness). To distinguish a solute diuresis from a water diuresis and to determine whether the diuresis is appropriate for the clinical circumstances, urine osmolality is measured. The average person excretes between 600 and 800 mosmol of solutes per day, primarily as urea and electrolytes. If the urine output is >3 L/d and the urine is dilute (<250 mosmol/L), total osmolar excretion is normal and a water diuresis is present. This circumstance could arise from polydipsia, inadequate secretion of AVP (*central diabetes*

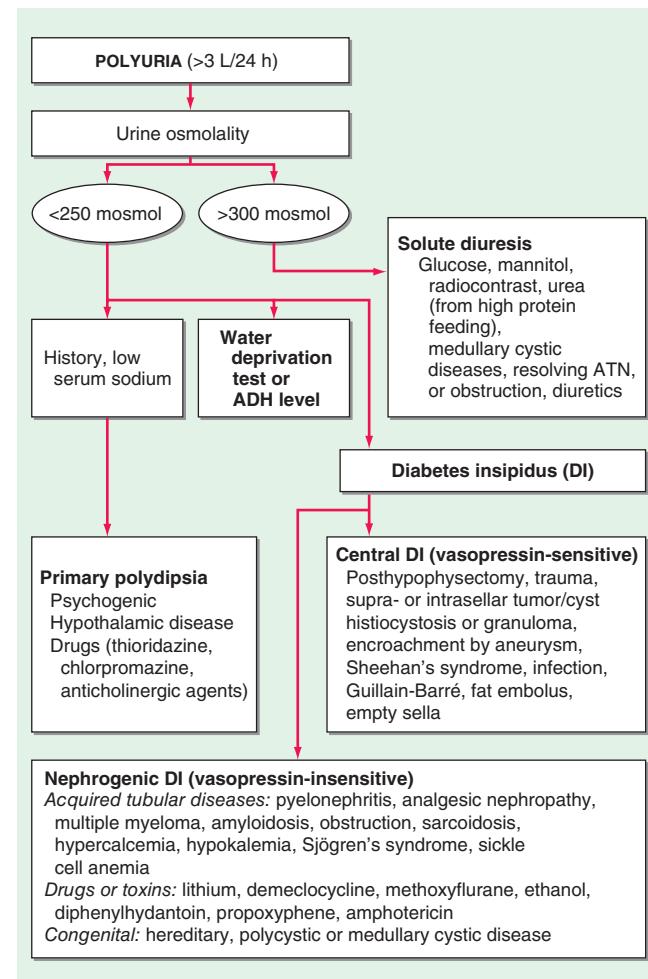


FIGURE 52-4 Approach to the patient with polyuria. ADH, antidiuretic hormone; ATN, acute tubular necrosis.

insipidus), or failure of renal tubules to respond to AVP (*nephrogenic diabetes insipidus*). If the urine volume is >3 L/d and urine osmolality is >300 mosmol/L, a solute diuresis is clearly present and a search for the responsible solute(s) is mandatory.

Excessive filtration of a poorly reabsorbed solute such as glucose or mannitol can depress reabsorption of NaCl and water in the proximal tubule and lead to enhanced excretion in the urine. Poorly controlled diabetes mellitus with glucosuria is the most common cause of a solute diuresis, leading to volume depletion and serum hypertonicity. Since the urine sodium concentration is less than that of blood, more water than sodium is lost, causing hypernatremia and hypertonicity. Common iatrogenic solute diuresis occurs in association with mannitol administration, radiocontrast media, and high-protein feedings (enteral or parenteral), leading to increased urea production and excretion. Less commonly, excessive sodium loss may result from cystic renal diseases or Bartter's syndrome or may develop during a tubulointerstitial process (such as resolving ATN). In these so-called salt-wasting disorders, the tubule damage results in direct impairment of sodium reabsorption and indirectly reduces the responsiveness of the tubule to aldosterone. Usually, the sodium losses are mild, and the obligatory urine output is <2 L/d; resolving ATN and postobstructive diuresis are exceptions and may be associated with significant natriuresis and polyuria.

Formation of large volumes of dilute urine is usually due to polydipsic states or diabetes insipidus. Primary polydipsia can result from habit, psychiatric disorders, neurologic lesions, or medications. During deliberate polydipsia, extracellular fluid volume is normal or expanded and plasma AVP levels are reduced because serum osmolality tends to be near the lower limits of normal. Urine osmolality is also maximally dilute at 50 mosmol/L.

Central diabetes insipidus may be idiopathic in origin or secondary to a variety of conditions, including hypophysectomy, trauma, neoplastic, inflammatory, vascular, or infectious hypothalamic diseases. Idiopathic central diabetes insipidus is associated with selective destruction of the AVP-secreting neurons in the supraoptic and paraventricular nuclei and can either be inherited as an autosomal dominant trait or occur spontaneously. Nephrogenic diabetes insipidus can occur in a variety of clinical situations, as summarized in Fig. 52-4.

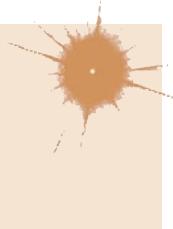
A plasma AVP level is recommended as the best method for distinguishing between central and nephrogenic diabetes insipidus. Assays for circulating copeptin, a peptide that is cleaved from pre-pro-AVP during axonal transport in the posterior pituitary, are also now available in many centers. A water deprivation test plus exogenous AVP may distinguish primary polydipsia from central and nephrogenic diabetes insipidus. Measurement of hypertonic saline-stimulated plasma copeptin, if available, can substitute for water deprivation testing. **For a detailed discussion, see Chap. 381.**

ACKNOWLEDGMENT

Julie Lin and Brad Denker contributed to this chapter in the 19th edition, and some material from that chapter has been retained here.

FURTHER READING

- EMMETT M et al: Approach to the patient with kidney disease, in *Brenner and Rector's The Kidney*, 10th ed, K Skorecki et al (eds). Philadelphia, W.B. Saunders & Company, 2016, pp. 754–779.
- ENEANYA ND et al: Reconsidering the consequences of using race to estimate kidney function. *JAMA* 322:113, 2019.
- KÖHLER H et al: Acanthocyturia—a characteristic marker for glomerular bleeding. *Kidney Int* 40:115, 1991.
- PERAZELLA MA: The urine sediment as a biomarker of kidney disease. *Am J Kidney Dis* 66:748, 2015.
- POWE NR: Black kidney function matters: Use or misuse of race? *JAMA* 324:737, 2020.
- WEISORD SD et al: Prevention and management of acute kidney injury in *Brenner and Rector's The Kidney*, 11th ed, ASL Yu et al: (eds). Philadelphia, W.B. Saunders & Company, 2020, pp. 940–977.



SODIUM AND WATER

COMPOSITION OF BODY FLUIDS

Water is the most abundant constituent in the body, comprising ~50% of body weight in women and 60% in men. Total-body water is distributed in two major compartments: 55–75% is intracellular (intracellular fluid [ICF]), and 25–45% is extracellular (extracellular fluid [ECF]). The ECF is further subdivided into intravascular (plasma water) and extravascular (interstitial) spaces in a ratio of 1:3. Fluid movement between the intravascular and interstitial spaces occurs across the capillary wall and is determined by Starling forces, i.e., capillary hydraulic pressure and colloid osmotic pressure. The transcapillary hydraulic pressure gradient exceeds the corresponding oncotic pressure gradient, thereby favoring the movement of plasma ultrafiltrate into the extravascular space. The return of fluid into the intravascular compartment occurs via lymphatic flow.

The solute or particle concentration of a fluid is known as its osmolality, expressed as milliosmoles per kilogram of water (mOsm/kg). Water easily diffuses across most cell membranes to achieve osmotic equilibrium (ECF osmolality = ICF osmolality). Notably, the extracellular and intracellular solute compositions differ considerably owing to the activity of various transporters, channels, and ATP-driven membrane pumps. The major ECF particles are Na^+ and its accompanying anions Cl^- and HCO_3^- , whereas K^+ and organic phosphate esters (ATP, creatine phosphate, and phospholipids) are the predominant ICF osmoles. Solutes that are restricted to the ECF or the ICF determine the “tonicity” or effective osmolality of that compartment. Certain solutes, particularly urea, do not contribute to water shifts across most membranes and are thus known as *ineffective osmoles*.

Water Balance Vasopressin secretion, water ingestion, and renal water transport collaborate to maintain human body fluid osmolality between 280 and 295 mOsm/kg. Vasopressin (AVP) is synthesized in magnocellular neurons within the hypothalamus; the distal axons of these neurons project to the posterior pituitary or neurohypophysis, from which AVP is released into the circulation. A network of central “osmoreceptor” neurons, which includes the AVP-expressing magnocellular neurons themselves, sense circulating osmolality via nonselective, stretch-activated cation channels. These osmoreceptor neurons are activated or inhibited by modest increases and decreases in circulating osmolality, respectively; activation leads to AVP release and thirst.

AVP secretion is stimulated as systemic osmolality increases above a threshold level of ~285 mOsm/kg, above which there is a linear relationship between osmolality and circulating AVP (Fig. 53-1). Thirst and thus water ingestion are also activated at ~285 mOsm/kg, beyond which there is an equivalent linear increase in the perceived intensity of thirst as a function of circulating osmolality. Changes in blood volume and blood pressure are also direct stimuli for AVP release and thirst, albeit with a less sensitive response profile. Of perhaps greater clinical relevance to the pathophysiology of water homeostasis, ECF volume strongly modulates the relationship between circulating osmolality and AVP release, such that hypovolemia reduces the osmotic threshold and increases the slope of the response curve to osmolality; *hypervolemia* has an opposite effect, increasing the osmotic threshold and reducing the slope of the response curve (Fig. 53-1). Notably, AVP has a half-life in the circulation of only 10–20 min; thus, changes in ECF volume and/or circulating osmolality can rapidly affect water homeostasis. In addition to volume status, a number of other “nonosmotic” stimuli have potent activating effects on osmosensitive neurons and AVP

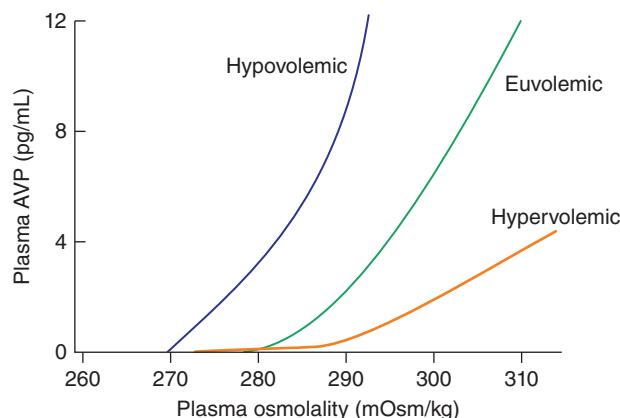


FIGURE 53-1 Circulating levels of vasopressin (AVP) in response to changes in osmolality. Plasma AVP becomes detectable in euvolemic, healthy individuals at a threshold of ~285 mOsm/kg, above which there is a linear relationship between osmolality and circulating AVP. The AVP response to osmolality is modulated strongly by volume status. The osmotic threshold is thus slightly lower in hypovolemia, with a steeper response curve; hypovolemia reduces the sensitivity of circulating AVP levels to osmolality.

release, including nausea, intracerebral angiotensin II, serotonin, and multiple drugs.

The excretion or retention of electrolyte-free water by the kidney is modulated by circulating AVP. AVP acts on renal, V₂-type receptors in the thick ascending limb of Henle and principal cells of the collecting duct (CD), increasing intracellular levels of cyclic AMP and activating protein kinase A (PKA)-dependent phosphorylation of multiple transport proteins. The AVP- and PKA-dependent activation of Na⁺-Cl⁻ and K⁺ transport by the thick ascending limb of the loop of Henle (TALH) is a key participant in the countercurrent mechanism (Fig. 53-2). The countercurrent mechanism ultimately increases the interstitial osmolality in the inner medulla of the kidney, driving water absorption across the renal CD. However, water, salt, and solute transport by both proximal and distal nephron segments participates in the renal concentrating mechanism (Fig. 53-2). Water transport across apical and basolateral aquaporin-1 water channels in the descending thin limb of the loop of Henle is thus involved, as is passive absorption of Na⁺-Cl⁻ by

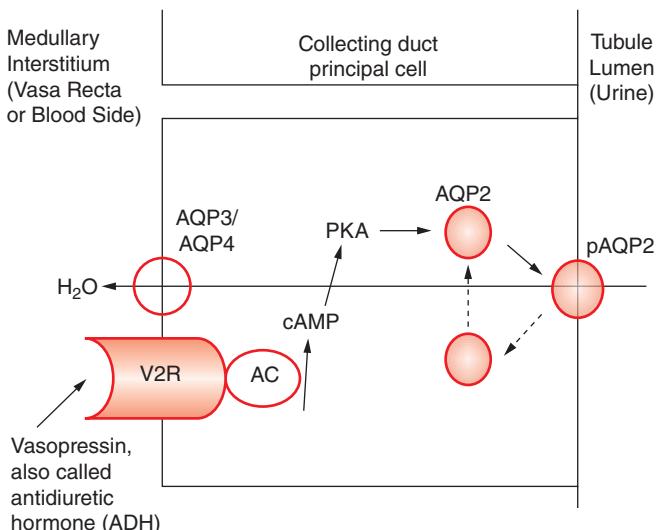


FIGURE 53-3 Vasopressin and the regulation of water permeability in the renal collecting duct. Vasopressin binds to the type 2 vasopressin receptor (V2R) on the basolateral membrane of principal cells, activates adenylyl cyclase (AC), increases intracellular cyclic adenosine monophosphate (cAMP), and stimulates protein kinase A (PKA) activity. Cytoplasmic vesicles carrying aquaporin-2 (AQP) water channel proteins are inserted into the luminal membrane in response to vasopressin, thereby increasing the water permeability of this membrane. When vasopressin stimulation ends, water channels are retrieved by an endocytic process and water permeability returns to its low basal rate. The AQP3 and AQP4 water channels are expressed on the basolateral membrane and complete the transcellular pathway for water reabsorption. pAQP2, phosphorylated aquaporin-2. (From Annals of Internal Medicine JM Sands, DG Bichet: Nephrogenic diabetes insipidus. 144(3):186, 2006. Copyright © 2006 American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.)

the thin ascending limb, via apical and basolateral CLC-K1 chloride channels and paracellular Na⁺ transport. Renal urea transport in turn plays important roles in the generation of the medullary osmotic gradient and the ability to excrete solute-free water under conditions of both high and low protein intake (Fig. 53-2).

AVP-induced, PKA-dependent phosphorylation of the aquaporin-2 water channel in principal cells stimulates the insertion of active water channels into the lumen of the CD, resulting in transepithelial water absorption down the medullary osmotic gradient (Fig. 53-3). Under “antidiuretic” conditions, with increased circulating AVP, the kidney reabsorbs water filtered by the glomerulus, equilibrating the osmolality across the CD epithelium to excrete a hypertonic, “concentrated” urine (osmolality of up to 1200 mOsm/kg). In the absence of circulating AVP, insertion of aquaporin-2 channels and water absorption across the CD is essentially abolished, resulting in secretion of a hypotonic, dilute urine (osmolality as low as 30–50 mOsm/kg). Abnormalities in this “final common pathway” are involved in most disorders of water homeostasis, e.g., a reduced or absent insertion of active aquaporin-2 water channels into the membrane of principal cells in diabetes insipidus (DI).

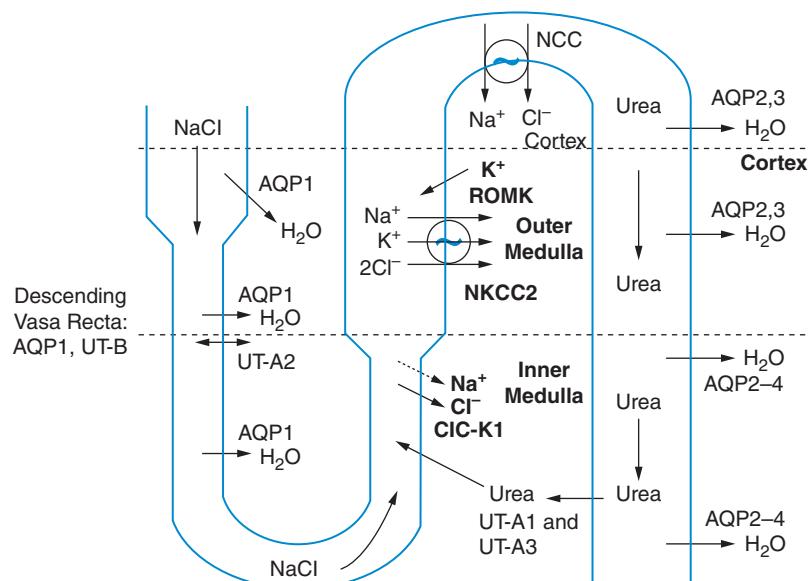


FIGURE 53-2 The renal concentrating mechanism. Water, salt, and solute transport by both proximal and distal nephron segments participates in the renal concentrating mechanism (see text for details). Diagram showing the location of the major transport proteins involved; a loop of Henle is depicted on the left, collecting duct on the right. AQP, aquaporin; CLC-K1, chloride channel; NKCC2, Na⁺-K⁺-2Cl⁻ cotransporter; ROMK, renal outer medullary K⁺ channel; UT, urea transporter. (Republished with permission of American Society of Nephrology, from Molecular approaches to urea transporters, JM Sands, 13(11), 2002; permission conveyed through Copyright Clearance Center, Inc.)

Maintenance of Arterial Circulatory Integrity
Sodium is actively pumped out of cells by the Na⁺/K⁺-ATPase membrane pump. In consequence, 85–90% of body Na⁺ is extracellular, and the ECF volume (ECFV) is a function of total-body Na⁺ content. Arterial perfusion and circulatory integrity are, in turn, determined by renal Na⁺ retention or excretion, in addition to the modulation of systemic arterial resistance. Within the kidney, Na⁺ is filtered by the glomeruli and then sequentially reabsorbed by the renal tubules. The Na⁺ cation is typically reabsorbed with the chloride anion (Cl⁻), and thus, chloride homeostasis also affects the ECFV. On a quantitative level, at a glomerular filtration rate (GFR) of 180 L/d and

serum Na^+ of ~140 mM, the kidney filters some 25,200 mmol/d of Na^+ . This is equivalent to ~1.5 kg of salt, which would occupy roughly 10 times the extracellular space; 99.6% of filtered $\text{Na}^+ \text{-Cl}^-$ must be reabsorbed to excrete 100 mM per day. Minute changes in renal $\text{Na}^+ \text{-Cl}^-$ excretion will thus have significant effects on the ECFV, leading to edema syndromes or hypovolemia.

Approximately two-thirds of filtered $\text{Na}^+ \text{-Cl}^-$ is reabsorbed by the renal proximal tubule, via both paracellular and transcellular mechanisms. The TALH subsequently reabsorbs another 25–30% of filtered $\text{Na}^+ \text{-Cl}^-$ via the apical, furosemide-sensitive $\text{Na}^+ \text{-K}^+ \text{-2Cl}^-$ cotransporter. The adjacent aldosterone-sensitive distal nephron, comprising the distal convoluted tubule (DCT), connecting tubule (CNT), and CD, accomplishes the “fine-tuning” of renal $\text{Na}^+ \text{-Cl}^-$ excretion. The thiazide-sensitive apical $\text{Na}^+ \text{-Cl}^-$ cotransporter (NCC) reabsorbs 5–10% of filtered $\text{Na}^+ \text{-Cl}^-$ in the DCT. Principal cells in the CNT and CD reabsorb Na^+ via electrogenic, amiloride-sensitive epithelial Na^+ channels (ENaC); Cl^- ions are primarily reabsorbed by adjacent intercalated cells, via apical Cl^- exchange ($\text{Cl}^- \text{-OH}^-$ and $\text{Cl}^- \text{-HCO}_3^-$ exchange, mediated by the SLC26A4 anion exchanger) (Fig. 53-4).

Renal tubular reabsorption of filtered $\text{Na}^+ \text{-Cl}^-$ is regulated by multiple circulating and paracrine hormones, in addition to the activity of renal nerves. Angiotensin II activates proximal $\text{Na}^+ \text{-Cl}^-$ reabsorption, as do adrenergic receptors under the influence of renal sympathetic innervation; locally generated dopamine, in contrast, has a *natriuretic* effect. Aldosterone primarily activates $\text{Na}^+ \text{-Cl}^-$ reabsorption within the aldosterone-sensitive distal nephron. In particular, aldosterone activates the ENaC channel in principal cells, inducing Na^+ absorption and promoting K^+ excretion (Fig. 53-4).

Circulatory integrity is critical for the perfusion and function of vital organs. “Underfilling” of the arterial circulation is sensed by ventricular and vascular pressure receptors, resulting in a neurohumoral activation

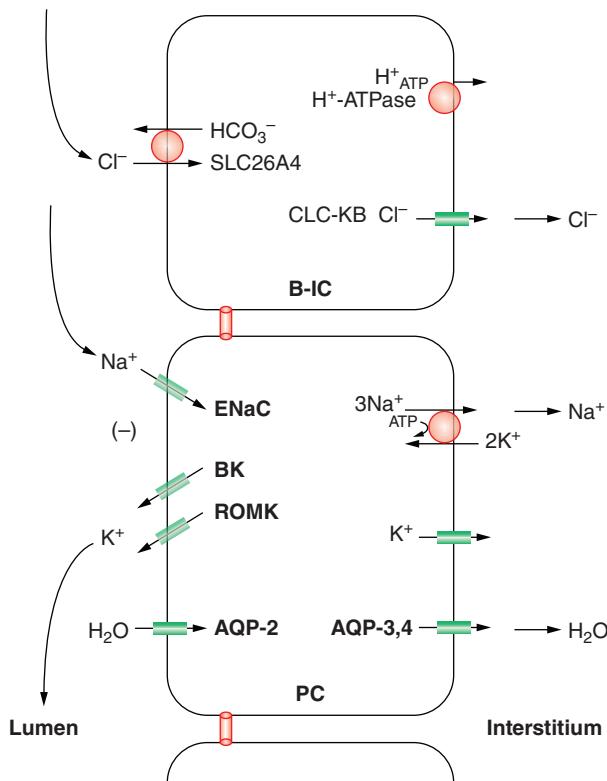


FIGURE 53-4 Sodium, water, and potassium transport in principal cells (PC) and adjacent β -intercalated cells (B-IC). The absorption of Na^+ via the amiloride-sensitive epithelial sodium channel (ENaC) generates a lumen-negative potential difference, which drives K^+ excretion through the apical secretory K^+ channel ROMK (renal outer medullary K^+ channel) and/or the flow-dependent BK channel. Transepithelial Cl^- transport occurs in adjacent β -intercalated cells, via apical $\text{Cl}^- \text{-HCO}_3^-$ and $\text{Cl}^- \text{-OH}^-$ exchange (SLC26A4 anion exchanger, also known as pendrin) basolateral CLC chloride channels. Water is absorbed down the osmotic gradient by principal cells, through the apical aquaporin-2 (AQP-2) and basolateral aquaporin-3 and aquaporin-4 (Fig. 53-3).

(increased sympathetic tone, activation of the renin-angiotensin-aldosterone axis, and increased circulating AVP) that synergistically increases renal $\text{Na}^+ \text{-Cl}^-$ reabsorption, vascular resistance, and renal water reabsorption. This occurs in the context of decreased cardiac output, as occurs in hypovolemic states, low-output cardiac failure, decreased oncotic pressure, and/or increased capillary permeability. Alternatively, excessive arterial vasodilation results in *relative* arterial underfilling, leading to neurohumoral activation in the defense of tissue perfusion. These physiologic responses play important roles in many of the disorders discussed in this chapter. In particular, it is important to appreciate that AVP functions in the defense of circulatory integrity, inducing vasoconstriction, increasing sympathetic nervous system tone, increasing renal retention of both water and $\text{Na}^+ \text{-Cl}^-$, and modulating the arterial baroreceptor reflex. Most of these responses involve activation of systemic $\text{V}_{1\alpha}$ AVP receptors, but concomitant activation of V_2 receptors in the kidney can result in renal water retention and hyponatremia.

HYPVOLEMIA

Etiology True volume depletion, or hypovolemia, generally refers to a state of combined salt and water loss, leading to contraction of the ECFV. The loss of salt and water may be renal or nonrenal in origin.

RENAL CAUSES Excessive urinary $\text{Na}^+ \text{-Cl}^-$ and water loss is a feature of several conditions. A high filtered load of endogenous solutes, such as glucose and urea, can impair tubular reabsorption of $\text{Na}^+ \text{-Cl}^-$ and water, leading to an osmotic diuresis. Exogenous mannitol, often used to decrease intracerebral pressure, is filtered by glomeruli but not reabsorbed by the proximal tubule, thus causing an osmotic diuresis. Pharmacologic diuretics selectively impair $\text{Na}^+ \text{-Cl}^-$ reabsorption at specific sites along the nephron, leading to increased urinary $\text{Na}^+ \text{-Cl}^-$ excretion. Other drugs can induce natriuresis as a side effect. For example, acetazolamide can inhibit proximal tubular $\text{Na}^+ \text{-Cl}^-$ absorption via its inhibition of carbonic anhydrase; other drugs, such as the antibiotics trimethoprim (TMP) and pentamidine, inhibit distal tubular Na^+ reabsorption through the amiloride-sensitive ENaC channel, leading to urinary $\text{Na}^+ \text{-Cl}^-$ loss. Hereditary defects in renal transport proteins are also associated with reduced reabsorption of filtered $\text{Na}^+ \text{-Cl}^-$ and/or water. Alternatively, mineralocorticoid deficiency, mineralocorticoid resistance, or inhibition of the mineralocorticoid receptor (MLR) can reduce $\text{Na}^+ \text{-Cl}^-$ reabsorption by the aldosterone-sensitive distal nephron. Finally, tubulointerstitial injury, as occurs in interstitial nephritis, acute tubular injury, or obstructive uropathy, can reduce distal tubular $\text{Na}^+ \text{-Cl}^-$ and/or water absorption.

Excessive excretion of free water, i.e., water without electrolytes, can also lead to hypovolemia. However, the effect on ECFV is usually less marked, given that two-thirds of the water volume is lost from the ICF. Excessive renal water excretion occurs in the setting of decreased circulating AVP or renal resistance to AVP (central and nephrogenic DI, respectively).

EXTRARENAL CAUSES Nonrenal causes of hypovolemia include fluid loss from the gastrointestinal tract, skin, and respiratory system. Accumulations of fluid within specific tissue compartments, typically the interstitium, peritoneum, or gastrointestinal tract, can also cause hypovolemia.

Approximately 9 L of fluid enter the gastrointestinal tract daily, 2 L by ingestion and 7 L by secretion; almost 98% of this volume is absorbed, such that daily fecal fluid loss is only 100–200 mL. Impaired gastrointestinal reabsorption or enhanced secretion of fluid can cause hypovolemia. Because gastric secretions have a low pH (high H^+ concentration), whereas biliary, pancreatic, and intestinal secretions are alkaline (high HCO_3^- concentration), vomiting and diarrhea are often accompanied by metabolic alkalosis and acidosis, respectively.

Evaporation of water from the skin and respiratory tract (so-called “insensible losses”) constitutes the major route for loss of solute-free water, which is typically 500–650 mL/d in healthy adults. This evaporative loss can increase during febrile illness or prolonged heat exposure. Hyperventilation can also increase insensible losses via the respiratory tract, particularly in ventilated patients; the humidity of inspired air

is another determining factor. In addition, increased exertion and/or ambient temperature will increase insensible losses via sweat, which is hypotonic to plasma. Profuse sweating without adequate repletion of water and $\text{Na}^+ \text{-Cl}^-$ can thus lead to both hypovolemia and hypertonicity. Alternatively, replacement of these insensible losses with a surfeit of free water, without adequate replacement of electrolytes, may lead to hypovolemic hyponatremia.

Excessive fluid accumulation in interstitial and/or peritoneal spaces can also cause intravascular hypovolemia. Increases in vascular permeability and/or a reduction in oncotic pressure (hypoalbuminemia) alter Starling forces, resulting in excessive “third spacing” of the ECFV. This occurs in sepsis syndrome, burns, pancreatitis, nutritional hypoalbuminemia, and peritonitis. Alternatively, distributive hypovolemia can occur due to accumulation of fluid within specific compartments, for example, within the bowel lumen in gastrointestinal obstruction or ileus. Hypovolemia can also occur after extracorporeal hemorrhage or after significant hemorrhage into an expandable space, for example, the retroperitoneum.

Diagnostic Evaluation A careful history will usually determine the etiologic cause of hypovolemia. Symptoms of hypovolemia are non-specific and include fatigue, weakness, thirst, and postural dizziness; more severe symptoms and signs include oliguria, cyanosis, abdominal and chest pain, and confusion or obtundation. Associated electrolyte disorders may cause additional symptoms, for example, muscle weakness in patients with hypokalemia. On examination, diminished skin turgor and dry oral mucous membranes are less than ideal markers of a decreased ECFV in adult patients; more reliable signs of hypovolemia include a decreased jugular venous pressure (JVP), orthostatic tachycardia (an increase of >15–20 beats/min upon standing), and orthostatic hypotension (a >10–20 mmHg drop in blood pressure on standing). More severe fluid loss leads to hypovolemic shock, with hypotension, tachycardia, peripheral vasoconstriction, and peripheral hypoperfusion; these patients may exhibit peripheral cyanosis, cold extremities, oliguria, and altered mental status.

Routine chemistries may reveal an increase in blood urea nitrogen (BUN) and creatinine, reflective of a decrease in GFR. Creatinine is the more dependable measure of GFR, because BUN levels may be influenced by an increase in tubular reabsorption (“prerenal azotemia”), an increase in urea generation in catabolic states, hyperalimentation, or gastrointestinal bleeding, and/or a decreased urea generation in decreased protein intake. In hypovolemic shock, liver function tests and cardiac biomarkers may show evidence of hepatic and cardiac ischemia, respectively. Routine chemistries and/or blood gases may reveal evidence of acid-base disorders. For example, bicarbonate loss due to diarrheal illness is a very common cause of metabolic acidosis; alternatively, patients with severe hypovolemic shock may develop lactic acidosis with an elevated anion gap.

The neurohumoral response to hypovolemia stimulates an increase in renal tubular Na^+ and water reabsorption. Therefore, the urine Na^+ concentration is typically <20 mM in nonrenal causes of hypovolemia, with a urine osmolality of >450 mOsm/kg. The reduction in both GFR and distal tubular Na^+ delivery may cause a defect in renal potassium excretion, with an increase in plasma K^+ concentration. Of note, patients with hypovolemia and a hypochloremic alkalosis due to vomiting, diarrhea, or diuretics will typically have a urine Na^+ concentration >20 mM and urine pH of >7.0, due to the increase in filtered HCO_3^- ; the urine Cl^- concentration in this setting is a more accurate indicator of volume status, with a level <25 mM suggestive of hypovolemia. The urine Na^+ concentration is often >20 mM in patients with *renal* causes of hypovolemia, such as acute tubular necrosis; similarly, patients with DI will have an inappropriately dilute urine.

TREATMENT

Hypovolemia

The therapeutic goals in hypovolemia are to restore normovolemia and replace ongoing fluid losses. Mild hypovolemia can usually be treated with oral hydration and resumption of a normal

maintenance diet. More severe hypovolemia requires intravenous hydration, tailoring the choice of solution to the underlying pathophysiology. Isotonic, “normal” saline (0.9% NaCl, 154 mM Na^+) is the most appropriate resuscitation fluid for normonatremic or hyponatremic patients with severe hypovolemia; colloid solutions such as intravenous albumin are not demonstrably superior for this purpose. Hypernatremic patients should receive a hypotonic solution, 5% dextrose if there has only been water loss (as in DI), or hypotonic saline (1/2 or 1/4 normal saline) if there has been water and $\text{Na}^+ \text{-Cl}^-$ loss; changes in free water administration should be made if necessary, based on frequent measuring of serum chemistries. Patients with bicarbonate loss and metabolic acidosis, as occur frequently in diarrhea, should receive intravenous bicarbonate, either an isotonic solution (150 meq of $\text{Na}^+ \text{-HCO}_3^-$ in 5% dextrose) or a more hypotonic bicarbonate solution in dextrose or dilute saline. Patients with severe hemorrhage or anemia should receive red cell transfusions, without increasing the hematocrit beyond 35%.

SODIUM DISORDERS

Disorders of serum Na^+ concentration are caused by abnormalities in water homeostasis, leading to changes in the relative ratio of Na^+ to body water. Water intake and circulating AVP constitute the two key effectors in the defense of serum osmolality; defects in one or both of these two defense mechanisms cause most cases of hyponatremia and hypernatremia. In contrast, abnormalities in sodium homeostasis per se lead to a deficit or surplus of whole-body $\text{Na}^+ \text{-Cl}^-$ content, a key determinant of the ECFV and circulatory integrity. Notably, volume status also modulates the release of AVP by the posterior pituitary, such that hypovolemia is associated with higher circulating levels of the hormone at each level of serum osmolality. Similarly, in “hypervolemic” causes of arterial underfilling, e.g., heart failure and cirrhosis, the associated neurohumoral activation encompasses an increase in circulating AVP, leading to water retention and hyponatremia. Therefore, a key concept in sodium disorders is that the absolute plasma Na^+ concentration tells one nothing about the volume status of a given patient, which furthermore must be taken into account in the diagnostic and therapeutic approach.

HYPONATREMIA

Hyponatremia, which is defined as a plasma Na^+ concentration <135 mM, is a very common disorder, occurring in up to 22% of hospitalized patients. This disorder is almost always the result of an increase in circulating AVP and/or increased renal sensitivity to AVP, combined with an intake of free water; a notable exception is hyponatremia due to low solute intake (see below). The underlying pathophysiology for the exaggerated or “inappropriate” AVP response differs in patients with hyponatremia as a function of their ECFV. Hyponatremia is thus subdivided diagnostically into three groups, depending on clinical history and volume status, i.e., “hypovolemic,” “euvolemic,” and “hypervolemic” (Fig. 53-5).

Hypovolemic Hyponatremia Hypovolemia causes a marked neurohumoral activation, increasing circulating levels of AVP. The increase in circulating AVP helps preserve blood pressure via vascular and baroreceptor V_{1A} receptors and increases water reabsorption via renal V_2 receptors; activation of V_2 receptors can lead to hyponatremia in the setting of increased free water intake. Nonrenal causes of hypovolemic hyponatremia include gastrointestinal loss (e.g., vomiting, diarrhea, tube drainage) and insensible loss (sweating, burns) of $\text{Na}^+ \text{-Cl}^-$ and water, in the absence of adequate oral replacement; urine Na^+ concentration is typically <20 mM. Notably, these patients may be clinically classified as euvolemic, with only the reduced urinary Na^+ concentration to indicate the cause of their hyponatremia. Indeed, a urine Na^+ concentration <20 mM, in the absence of a cause of hypervolemic hyponatremia, predicts a rapid increase in plasma Na^+ concentration in response to intravenous normal saline; saline therapy thus induces a water diuresis in this setting, as circulating AVP levels plummet.

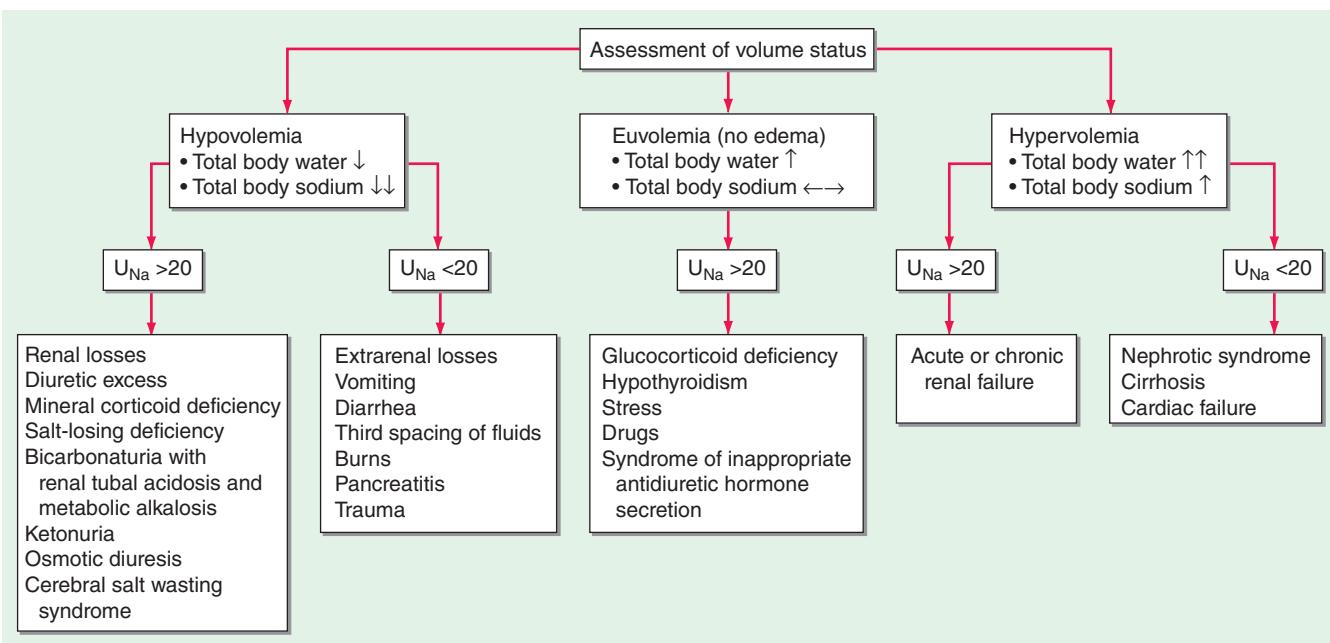


FIGURE 53-5 The diagnostic approach to hyponatremia. (Reproduced with permission from S Kumar, T Berl: Diseases of water metabolism, in RW Schrier [ed], *Atlas of Diseases of the Kidney*, Philadelphia, Current Medicine, Inc, 1999.)

The *renal* causes of hypovolemic hyponatremia share an inappropriate loss of Na^+-Cl^- in the urine, leading to volume depletion and an increase in circulating AVP; urine Na^+ concentration is typically $>20 \text{ mM}$ (Fig. 53-5). A deficiency in circulating aldosterone and/or its renal effects can lead to hyponatremia in primary adrenal insufficiency and other causes of hypoaldosteronism; hyperkalemia and hyponatremia in a hypotensive and/or hypovolemic patient with high urine Na^+ concentration (much greater than 20 mM) should strongly suggest this diagnosis. Salt-losing nephropathies may lead to hyponatremia when sodium intake is reduced, due to impaired renal tubular function; typical causes include reflux nephropathy, interstitial nephropathies, postobstructive uropathy, medullary cystic disease, and the recovery phase of acute tubular necrosis. Thiazide diuretics cause hyponatremia via a number of mechanisms, including polydipsia and diuretic-induced volume depletion. Notably, thiazides do not inhibit the renal concentrating mechanism, such that circulating AVP retains a full effect on renal water retention. In contrast, loop diuretics, which are less frequently associated with hyponatremia, inhibit Na^+-Cl^- and K^+ absorption by the TALH, blunting the countercurrent mechanism and reducing the ability to concentrate the urine. Increased excretion of an osmotically active nonreabsorbable or poorly reabsorbable solute can also lead to volume depletion and hyponatremia; important causes include glycosuria, ketonuria (e.g., in starvation or in diabetic or alcoholic ketoacidosis), and bicarbonaturia (e.g., in renal tubular acidosis or metabolic alkalosis, where the associated bicarbonaturia leads to loss of Na^+).

Finally, the syndrome of “cerebral salt wasting” is a rare cause of hypovolemic hyponatremia, encompassing hyponatremia with clinical hypovolemia and inappropriate natriuresis in association with intracranial disease; associated disorders include subarachnoid hemorrhage, traumatic brain injury, craniotomy, encephalitis, and meningitis. Distinction from the more common syndrome of inappropriate antidiuresis (SIAD) is critical because cerebral salt wasting will typically respond to aggressive Na^+-Cl^- repletion.

Hypervolemic Hyponatremia Patients with hypervolemic hyponatremia develop an increase in total-body Na^+-Cl^- that is accompanied by a proportionately *greater* increase in total-body water, leading to a reduced plasma Na^+ concentration. As in hypovolemic hyponatremia, the causative disorders can be separated by the effect on urine Na^+ concentration, with acute or chronic renal failure uniquely associated with an increase in urine Na^+ concentration (Fig. 53-5).

The pathophysiology of hyponatremia in the sodium-avid edematous disorders (congestive heart failure [CHF], cirrhosis, and nephrotic syndrome) is similar to that in hypovolemic hyponatremia, except that arterial filling and circulatory integrity is decreased due to the specific etiologic factors (e.g., cardiac dysfunction in CHF, peripheral vasodilation in cirrhosis). Urine Na^+ concentration is typically very low, i.e., $<10 \text{ mM}$, even after hydration with normal saline; this Na^+ -avid state may be obscured by diuretic therapy. The degree of hyponatremia provides an indirect index of the associated neurohumoral activation and is an important prognostic indicator in hypervolemic hyponatremia.

Euvolemic Hyponatremia Euvolemic hyponatremia can occur in moderate to severe hypothyroidism, with correction after achieving a euthyroid state. Severe hyponatremia can also be a consequence of secondary adrenal insufficiency due to pituitary disease; whereas the deficit in circulating aldosterone in primary adrenal insufficiency causes *hypovolemic* hyponatremia, the predominant glucocorticoid deficiency in secondary adrenal failure is associated with *euvolemic* hyponatremia. Glucocorticoids exert a negative feedback on AVP release by the posterior pituitary such that hydrocortisone replacement in these patients can rapidly normalize the AVP response to osmolality, reducing circulating AVP.

The SIAD is the most frequent cause of euvolemic hyponatremia (**Table 53-1**). The generation of hyponatremia in SIAD requires an intake of free water, with persistent intake at serum osmolalities that are lower than the usual threshold for thirst; as one would expect, the osmotic threshold and osmotic response curves for the sensation of thirst are shifted downward in patients with SIAD. Four distinct patterns of AVP secretion have been recognized in patients with SIAD, independent for the most part of the underlying cause. Unregulated, erratic AVP secretion is seen in about a third of patients, with no obvious correlation between serum osmolality and circulating AVP levels. Other patients fail to suppress AVP secretion at lower serum osmolalities, with a normal response curve to hyperosmolar conditions; others have a “reset osmostat,” with a lower threshold osmolality and a left-shifted osmotic response curve. Finally, the fourth subset of patients have essentially no detectable circulating AVP, suggesting either a gain in function in renal water reabsorption or a circulating antidiuretic substance that is distinct from AVP. Gain-in-function mutations of a single specific residue in the V_2 AVP receptor have been described in some of these patients, leading to constitutive activation of the receptor in the absence of AVP and “nephrogenic” SIAD.

TABLE 53-1 Causes of the Syndrome of Inappropriate Antidiuresis (SIAD)

MALIGNANT DISEASES	PULMONARY DISORDERS	DISORDERS OF THE CENTRAL NERVOUS SYSTEM	DRUGS	OTHER CAUSES
Carcinoma	Infections	Infection	Drugs that stimulate release of AVP or enhance its action	Hereditary (gain-of-function mutations in the vasopressin V ₂ receptor)
Lung	Bacterial pneumonia	Encephalitis	Chlorpropamide	Idiopathic
Small cell	Viral pneumonia	Meningitis	SSRIs	Transient
Mesothelioma	Pulmonary abscess	Brain abscess	Tricyclic antidepressants	Endurance exercise
Oropharynx	Tuberculosis	Rocky Mountain spotted fever	Clofibrate	General anesthesia
Gastrointestinal tract	Aspergillosis	AIDS	Carbamazepine	Nausea
Stomach	Asthma	Bleeding and masses	Vincristine	Pain
Duodenum	Cystic fibrosis	Subdural hematoma	Nicotine	Stress
Pancreas	Respiratory failure associated with positive-pressure breathing	Subarachnoid hemorrhage	Narcotics	
Genitourinary tract		Cerebrovascular accident	Antipsychotic drugs	
Ureter		Brain tumors	Ifosfamide	
Bladder		Head trauma	Cyclophosphamide	
Prostate		Hydrocephalus	Nonsteroidal anti-inflammatory drugs	
Endometrium		Cavernous sinus thrombosis	MDMA ("Ecstasy", "Molly")	
Endocrine thymoma		Other	AVP analogues	
Lymphomas		Multiple sclerosis	Desmopressin	
Sarcomas		Guillain-Barré syndrome	Oxytocin	
Ewing's sarcoma		Shy-Drager syndrome	Vasopressin	
		Delirium tremens		
		Acute intermittent porphyria		

Abbreviations: AVP, vasopressin; MDMA, 3,4-methylenedioxymethamphetamine; SSRI, selective serotonin reuptake inhibitor.

Source: From DH Ellison, T Berl: The syndrome of inappropriate antidiuresis. N Engl J Med 356:2064, 2007. Copyright © 2007 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

Strictly speaking, patients with SIAD are not euvolemic but are subclinically volume-expanded, due to AVP-induced water and Na⁺-Cl⁻ retention; "AVP escape" mechanisms invoked by sustained increases in AVP serve to limit distal renal tubular transport, preserving a modestly hypervolemic steady state. Serum uric acid is often low (<4 mg/dL) in patients with SIAD, consistent with suppressed proximal tubular transport in the setting of increased distal tubular Na⁺-Cl⁻ and water transport; in contrast, patients with hypovolemic hyponatremia will often be hyperuricemic due to a shared activation of proximal tubular Na⁺-Cl⁻ and urate transport.

Common causes of SIAD include pulmonary disease (e.g., pneumonia, tuberculosis, pleural effusion) and central nervous system (CNS) diseases (e.g., tumor, subarachnoid hemorrhage, meningitis). SIAD also occurs with malignancies, most commonly with small-cell lung carcinoma (75% of malignancy-associated SIAD); ~10% of patients with this tumor will have a plasma Na⁺ concentration of <130 mM at presentation. SIAD is also a frequent complication of certain drugs, most commonly the selective serotonin reuptake inhibitors (SSRIs). Other drugs can potentiate the renal effect of AVP, without exerting direct effects on circulating AVP levels (Table 53-1).

Low Solute Intake and Hyponatremia Hyponatremia can occasionally occur in patients with a very low intake of dietary solutes. Classically, this occurs in alcoholics whose sole nutrient is beer, hence the diagnostic label of *beer potomania*; beer is very low in protein and salt content, containing only 1–2 mM of Na⁺. The syndrome has also been described in nonalcoholic patients with highly restricted solute intake due to nutrient-restricted diets, e.g., extreme vegetarian diets. Patients with hyponatremia due to low solute intake typically present with a very low urine osmolality (<100–200 mOsm/kg) with a urine Na⁺ concentration that is <10–20 mM. The fundamental abnormality is the inadequate dietary intake of solutes; the reduced urinary solute excretion limits water excretion such that hyponatremia ensues after relatively modest polydipsia. AVP levels have not been reported in patients with beer potomania but are expected to be suppressed or rapidly suppressible with saline hydration; this fits with the overly rapid correction in plasma Na⁺ concentration that can be seen with saline hydration. Resumption of a normal diet and/or saline hydration will also correct the causative deficit in urinary solute excretion, such

that patients with beer potomania typically correct their plasma Na⁺ concentration promptly after admission to the hospital.

Clinical Features of Hyponatremia Hyponatremia induces generalized cellular swelling, a consequence of water movement down the osmotic gradient from the hypotonic ECF to the ICF. The symptoms of hyponatremia are primarily neurologic, reflecting the development of cerebral edema within a rigid skull. The initial CNS response to acute hyponatremia is an increase in interstitial pressure, leading to shunting of ECF and solutes from the interstitial space into the cerebrospinal fluid and then on into the systemic circulation. This is accompanied by an efflux of the major intracellular ions, Na⁺, K⁺, and Cl⁻, from brain cells. Acute hyponatremic encephalopathy ensues when these volume regulatory mechanisms are overwhelmed by a rapid decrease in tonicity, resulting in acute cerebral edema. Early symptoms can include nausea, headache, and vomiting. However, severe complications can rapidly evolve, including seizure activity, brainstem herniation, coma, and death. A key complication of acute hyponatremia is normocapneic or hypercapneic respiratory failure; the associated hypoxia may amplify the neurologic injury. Normocapneic respiratory failure in this setting is typically due to noncardiogenic, "neurogenic" pulmonary edema, with a normal pulmonary capillary wedge pressure.

Acute symptomatic hyponatremia is a medical emergency, occurring in a number of specific settings (Table 53-2). Women, particularly

TABLE 53-2 Causes of Acute Hyponatremia

Iatrogenic
Postoperative: premenopausal women
Hypotonic fluids with cause of ↑ vasopressin
Glycine irrigation: TURP, uterine surgery
Colonoscopy preparation
Recent institution of thiazides
Polydipsia
MDMA ("ecstasy," "Molly") ingestion
Exercise induced
Multifactorial, e.g., thiazide and polydipsia

Abbreviations: MDMA, 3,4-methylenedioxymethamphetamine; TURP, transurethral resection of the prostate.

before menopause, are much more likely than men to develop encephalopathy and severe neurologic sequelae. Acute hyponatremia often has an iatrogenic component, e.g., when hypotonic intravenous fluids are given to postoperative patients with an increase in circulating AVP. Exercise-associated hyponatremia, an important clinical issue at marathons and other endurance events, has similarly been linked to both a “nonosmotic” increase in circulating AVP and excessive free water intake. The recreational drugs Molly and Ecstasy, which share an active ingredient (MDMA, 3,4-methylenedioxymethamphetamine), cause a rapid and potent induction of both thirst and AVP, leading to severe acute hyponatremia.

Persistent, chronic hyponatremia results in an efflux of organic osmolytes (creatinine, betaine, glutamate, myoinositol, and taurine) from brain cells; this response reduces intracellular osmolality and the osmotic gradient favoring water entry. This reduction in intracellular osmolytes is largely complete within 48 h, the time period that clinically defines chronic hyponatremia; this temporal definition has considerable relevance for the treatment of hyponatremia (see below). The cellular response to chronic hyponatremia does not fully protect patients from symptoms, which can include vomiting, nausea, confusion, and seizures, usually at plasma Na^+ concentration <125 mM. Even patients who are judged “asymptomatic” can manifest subtle gait and cognitive defects that reverse with correction of hyponatremia; notably, chronic “asymptomatic” hyponatremia increases the risk of falls. Chronic hyponatremia also increases the risk of bony fractures owing to the associated neurologic dysfunction and to a hyponatremia-associated reduction in bone density. Therefore, every attempt should be made to safely correct the plasma Na^+ concentration in patients with chronic hyponatremia, even in the absence of overt symptoms (see the section on treatment of hyponatremia below).

The management of chronic hyponatremia is complicated significantly by the asymmetry of the cellular response to correction of plasma Na^+ concentration. Specifically, the *reaccumulation* of organic osmolytes by brain cells is attenuated and delayed as osmolality increases after correction of hyponatremia, sometimes resulting in degenerative loss of oligodendrocytes and an osmotic demyelination syndrome (ODS). Overly rapid correction of hyponatremia (>8–10 mM in 24 h or 18 mM in 48 h) causes hypertonic stress in astrocytes within brain regions prone to ODS, leading to generalized protein ubiquitination and endoplasmic reticulum stress due to activation of the unfolded protein response; this is accompanied by apoptotic and autophagic cell death. Rapid correction of hyponatremia also causes a disruption in integrity of the blood-brain barrier, allowing the entry of immune mediators that may contribute to demyelination. The lesions of ODS classically affect the pons, a neuroanatomic structure wherein the delay in the reaccumulation of osmotic osmolytes is particularly pronounced; clinically, patients with central pontine myelinolysis can present 1 or more days after overcorrection of hyponatremia with paraparesis or quadripareisis, dysphagia, dysarthria, diplopia, a “locked-in syndrome,” and/or loss of consciousness. Other regions of the brain can also be involved in ODS, most commonly in association with lesions of the pons but occasionally in isolation; in order of frequency, the lesions of extrapontine myelinolysis can occur in the cerebellum, lateral geniculate body, thalamus, putamen, and cerebral cortex or subcortex. Clinical presentation of ODS can, therefore, vary as a function of the extent and localization of extrapontine myelinolysis, with the reported development of ataxia, mutism, parkinsonism, dystonia, and catatonia. Relowering of plasma Na^+ concentration after overly rapid correction can prevent or attenuate ODS (see the section on treatment of hyponatremia below). However, even appropriately slow correction can be associated with ODS, particularly in patients with additional risk factors; these include alcoholism, malnutrition, hypokalemia, and liver transplantation.

Diagnostic Evaluation of Hyponatremia Clinical assessment of hyponatremic patients should focus on the underlying cause; a detailed drug history is particularly crucial (Table 53-1). A careful clinical assessment of volume status is obligatory for the classical diagnostic approach to hyponatremia (Fig. 53-5). Hyponatremia is frequently

multifactorial, particularly when severe; clinical evaluation should consider *all* the possible causes for excessive circulating AVP, including volume status, drugs, and the presence of nausea and/or pain. Radiologic imaging may also be appropriate to assess whether patients have a pulmonary or CNS cause for hyponatremia. A screening chest x-ray may fail to detect a small-cell carcinoma of the lung; computed tomography (CT) scanning of the thorax should be considered in patients at high risk for this tumor (e.g., patients with a smoking history).

Laboratory investigation should include a measurement of serum osmolality to exclude pseudohyponatremia, which is defined as the coexistence of hyponatremia with a normal or increased plasma tonicity. Most clinical laboratories measure plasma Na^+ concentration by testing diluted samples with automated ion-sensitive electrodes, correcting for this dilution by assuming that plasma is 93% water. This correction factor can be inaccurate in patients with pseudohyponatremia due to extreme hyperlipidemia and/or hyperproteinemia, in whom serum lipid or protein makes up a greater percentage of plasma volume. The measured osmolality should also be converted to the effective osmolality (tonicity) by subtracting the measured concentration of urea (divided by 2.8, if in mg/dL); patients with hyponatremia have an effective osmolality of <275 mOsm/kg.

Elevated BUN and creatinine in routine chemistries can also indicate renal dysfunction as a potential cause of hyponatremia, whereas hyperkalemia may suggest adrenal insufficiency or hypoaldosteronism. Serum glucose should also be measured; plasma Na^+ concentration falls by ~1.6–2.4 mM for every 100-mg/dL increase in glucose, due to glucose-induced water efflux from cells; this “true” hyponatremia resolves after correction of hyperglycemia. Measurement of serum uric acid should also be performed; whereas patients with SIAD-type physiology will typically be hypouricemic (serum uric acid <4 mg/dL), volume-depleted patients will often be hyperuricemic. In the appropriate clinical setting, thyroid, adrenal, and pituitary function should also be tested; hypothyroidism and secondary adrenal failure due to pituitary insufficiency are important causes of euvolemic hyponatremia, whereas primary adrenal failure causes hypovolemic hyponatremia. A cosyntropin stimulation test is necessary to assess for primary adrenal insufficiency.

Urine electrolytes and osmolality are crucial tests in the initial evaluation of hyponatremia. A urine Na^+ concentration <20–30 mM is consistent with hypovolemic hyponatremia, in the clinical absence of a hypervolemic, Na^+ -avid syndrome such as CHF (Fig. 53-5). In contrast, patients with SIAD will typically excrete urine with an Na^+ concentration that is >30 mM. However, there can be substantial overlap in urine Na^+ concentration values in patients with SIAD and hypovolemic hyponatremia, particularly in the elderly; the ultimate “gold standard” for the diagnosis of hypovolemic hyponatremia is the demonstration that plasma Na^+ concentration corrects after hydration with normal saline. Patients with thiazide-associated hyponatremia may also present with higher than expected urine Na^+ concentration and other findings suggestive of SIAD; one should defer making a diagnosis of SIAD in these patients until 1–2 weeks after discontinuing the thiazide. A urine osmolality <100 mOsm/kg is suggestive of polydipsia; urine osmolality >400 mOsm/kg indicates that AVP excess is playing a more dominant role, whereas intermediate values are more consistent with multifactorial pathophysiology (e.g., AVP excess with a significant component of polydipsia). Patients with hyponatremia due to decreased solute intake (beer potomania) typically have urine Na^+ concentration <20 mM and urine osmolality in the range of <100 to the low 200s. Finally, the measurement of urine K^+ concentration is required to calculate the urine-to-plasma electrolyte ratio, which is useful to predict the response to fluid restriction (see the section on treatment of hyponatremia below).

TREATMENT

Hyponatremia

Three major considerations guide the therapy of hyponatremia. First, the presence and/or severity of symptoms determine the urgency and goals of therapy. Patients with acute hyponatremia

(Table 53-2) present with symptoms that can range from headache, nausea, and/or vomiting, to seizures, obtundation, and central herniation; patients with chronic hyponatremia, present for >48 h, are less likely to have severe symptoms. Second, patients with chronic hyponatremia are at risk for ODS if plasma Na⁺ concentration is corrected by >8–10 mM within the first 24 h and/or by >18 mM within the first 48 h. Third, the response to interventions such as hypertonic saline, isotonic saline, or AVP antagonists can be highly unpredictable, such that frequent monitoring of plasma Na⁺ concentration during corrective therapy is imperative.

Once the urgency in correcting the plasma Na⁺ concentration has been established and appropriate therapy instituted, the focus should be on treatment or withdrawal of the underlying cause. Patients with euvolemic hyponatremia due to SIAD, hypothyroidism, or secondary adrenal failure will respond to successful treatment of the underlying cause, with an increase in plasma Na⁺ concentration. However, not all causes of SIAD are immediately reversible, necessitating pharmacologic therapy to increase the plasma Na⁺ concentration (see below). Hypovolemic hyponatremia will respond to intravenous hydration with isotonic normal saline, with a rapid reduction in circulating AVP and a brisk water diuresis; it may be necessary to reduce the rate of correction if the history suggests that hyponatremia has been chronic, i.e., present for >48 h (see below). Hypervolemic hyponatremia due to CHF will often respond to improved therapy of the underlying cardiomyopathy, e.g., following the institution or intensification of angiotensin-converting enzyme (ACE) inhibition. Finally, patients with hyponatremia due to beer potomania and low solute intake will respond very rapidly to intravenous saline and the resumption of a normal diet. Notably, patients with beer potomania have a very high risk of developing ODS, due to the associated hypokalemia, alcoholism, malnutrition, and high risk of overcorrecting the plasma Na⁺ concentration.

Water deprivation has long been a cornerstone of the therapy of chronic hyponatremia. However, patients who are excreting minimal electrolyte-free water will require aggressive fluid restriction; this can be very difficult for patients with SIAD to tolerate, given that their thirst is also inappropriately stimulated. The urine-to-plasma electrolyte ratio (urinary [Na⁺] + [K⁺]/plasma [Na⁺]) can be exploited as a quick indicator of electrolyte-free water excretion (Table 53-3); patients with a ratio of >1 should be more aggressively restricted (<500 mL/d) if possible, those with a ratio of ~1 should be restricted to 500–700 mL/d, and those with a ratio <1 should be restricted to <1 L/d. In hypokalemic patients, potassium replacement will serve to increase plasma Na⁺ concentration, given that

the plasma Na⁺ concentration is a function of both exchangeable Na⁺ and exchangeable K⁺ divided by total-body water; a corollary is that aggressive repletion of K⁺ has the potential to overcorrect the plasma Na⁺ concentration even in the absence of hypertonic saline. Plasma Na⁺ concentration will also tend to respond to an increase in dietary solute intake, which increases the ability to excrete free water; this can be accomplished with oral salt tablets and with newly available, palatable preparations of oral urea.

Patients in whom therapy with fluid restriction, potassium replacement, and/or increased solute intake fails may merit pharmacologic therapy to increase their plasma Na⁺ concentration. Some patients with SIAD initially respond to combined therapy with oral furosemide, 20 mg twice a day (higher doses may be necessary in renal insufficiency), and oral salt tablets; furosemide serves to inhibit the renal countercurrent mechanism and blunt urinary concentrating ability, whereas the salt tablets counteract diuretic-associated natriuresis. The risk of hypokalemia and/or renal dysfunction limits enthusiasm for this approach, which requires careful titration of diuretic and salt tablets. Demeclocycline is a potent inhibitor of principal cells and can be used in patients whose Na levels do not increase in response to furosemide and salt tablets. However, this agent can be associated with a reduction in GFR, due to excessive natriuresis and/or direct renal toxicity; it should be avoided in cirrhotic patients in particular, who are at higher risk of nephrotoxicity due to drug accumulation. If available, palatable preparations of oral urea can also be used to manage SIAD, with comparable efficacy to AVP antagonists (vaptans); the increase in solute excretion with oral urea ingestion increases free water excretion, thus reducing the plasma Na⁺.

AVP antagonists (vaptans) are highly effective in SIAD and in hypervolemic hyponatremia due to heart failure or cirrhosis, reliably increasing plasma Na⁺ concentration due to their “aquaretic” effects (augmentation of free water clearance). Most of these agents specifically antagonize the V₂ AVP receptor; tolvaptan is currently the only oral V₂ antagonist to be approved by the U.S. Food and Drug Administration. Conivaptan, the only available intravenous vaptan, is a mixed V_{1A}/V₂ antagonist, with a modest risk of hypotension due to V_{1A} receptor inhibition. Therapy with vaptans must be initiated in a hospital setting, with a liberalization of fluid restriction (>2 L/d) and close monitoring of plasma Na⁺ concentration. Although approved for the management of all but hypovolemic hyponatremia and acute hyponatremia, the clinical indications are limited. Oral tolvaptan is perhaps most appropriate for the management of significant and persistent SIAD (e.g., in small-cell lung carcinoma) that has not responded to water restriction and/or oral furosemide and salt tablets. Abnormalities in liver function tests have been reported with chronic tolvaptan therapy; hence, the use of this agent should be restricted to <1–2 months.

Treatment of acute symptomatic hyponatremia should include hypertonic 3% saline (513 mM) to acutely increase plasma Na⁺ concentration by 1–2 mM/h to a total of 4–6 mM; this modest increase is typically sufficient to alleviate severe acute symptoms, after which corrective guidelines for chronic hyponatremia are appropriate (see below). A bolus of 100 mL of hypertonic saline is more effective than an infusion, rapidly improving both serum sodium and mental status. For ongoing infusions, a number of equations have been developed to estimate the required rate of hypertonic saline, which has an Na⁺-Cl⁻ concentration of 513 mM. The traditional approach is to calculate an Na⁺ deficit, where the Na⁺ deficit = 0.6 × body weight × (target plasma Na⁺ concentration – starting plasma Na⁺ concentration), followed by a calculation of the required rate. Regardless of the method used to determine the rate of administration, the increase in plasma Na⁺ concentration can be highly unpredictable during treatment with hypertonic saline, due to rapid changes in the underlying physiology; plasma Na⁺ concentration should be monitored every 2–4 h during treatment, with appropriate changes in therapy based on the observed rate of change. The administration of supplemental oxygen and ventilatory support is also critical in acute hyponatremia, in the event

TABLE 53-3 Management of Hyponatremia

Water Deficit

- Estimate total-body water (TBW): 50% of body weight in women and 60% in men
- Calculate free-water deficit: [(Na⁺ – 140)/140] × TBW
- Administer deficit over 48–72 h, without decrease in plasma Na⁺ concentration by >10 mM/24 h

Ongoing Water Losses

- Calculate free-water clearance, C_{eH₂O}:

$$C_{eH_2O} = V \times \left(1 - \frac{U_{Na} + U_K}{P_{Na}}\right)$$

where V is urinary volume, U_{Na} is urinary [Na⁺], U_K is urinary [K⁺], and P_{Na} is plasma [Na⁺]

Insensible Losses

- ~10 mL/kg per day; less if ventilated, more if febrile

Total

- Add components to determine water deficit and ongoing water loss; correct the water deficit over 48–72 h and replace daily water loss. Avoid correction of plasma [Na⁺] by >10 mM/d.

that patients develop acute pulmonary edema or hypercapneic respiratory failure. Intravenous loop diuretics will help treat acute pulmonary edema and will also increase free water excretion, by interfering with the renal countercurrent multiplication system. AVP antagonists do *not* have an approved role in the management of acute hyponatremia.

The rate of correction should be comparatively slow in *chronic* hyponatremia (<6–8 mM in the first 24 h and <6 mM each subsequent 24h), so as to avoid ODS; lower target rates are appropriate in patients at particular risk for ODS, such as alcoholics or hypokalemic patients. Overcorrection of the plasma Na⁺ concentration can occur when AVP levels rapidly normalize, for example, following the treatment of patients with chronic hypovolemic hyponatremia with intravenous saline or following glucocorticoid replacement of patients with hypopituitarism and secondary adrenal failure. Approximately 10% of patients treated with vaptans will overcorrect; the risk is increased if water intake is not liberalized. In the event that the plasma Na⁺ concentration overcorrects following therapy, be it with hypertonic saline, isotonic saline, or a vaptan, hyponatremia can be safely reinduced or stabilized by the administration of the AVP *agonist* desmopressin acetate (DDAVP) and/or the administration of free water, typically intravenous D₅W; the goal is to prevent or reverse the development of ODS. Alternatively, the treatment of patients with marked hyponatremia can be initiated with the twice-daily administration of DDAVP to maintain constant AVP bioactivity, combined with the administration of hypertonic saline to slowly correct the serum sodium in a more controlled fashion, thus reducing upfront the risk of overcorrection.

HYPERNATREMIA

Etiology Hypernatremia is defined as an increase in the plasma Na⁺ concentration to >145 mM. Considerably less common than hyponatremia, hypernatremia is nonetheless associated with mortality rates of as high as 40–60%, mostly due to the severity of the associated underlying disease processes. Hypernatremia is usually the result of a combined water and electrolyte deficit, with losses of H₂O in excess of Na⁺. Less frequently, the ingestion or iatrogenic administration of excess Na⁺ can be causative, for example, after IV administration of excessive hypertonic Na⁺-Cl⁻ or Na⁺-HCO₃⁻ (Fig. 53-6).

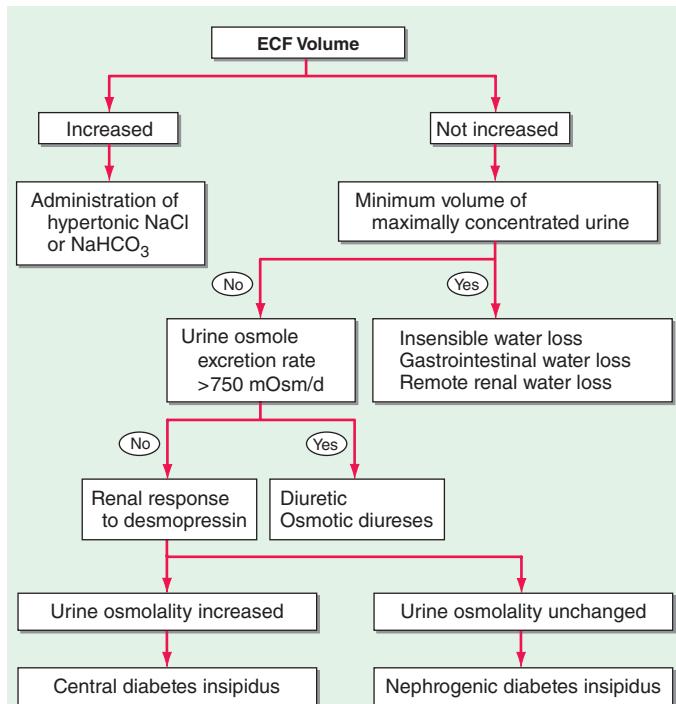


FIGURE 53-6 The diagnostic approach to hypernatremia. ECF, extracellular fluid.

Elderly individuals with reduced thirst and/or diminished access to fluids are at the highest risk of developing hypernatremia. Patients with hypernatremia may rarely have a central defect in hypothalamic osmoreceptor function, with a mixture of both decreased thirst and reduced AVP secretion. Causes of this adipsic DI include primary or metastatic tumor, occlusion or ligation of the anterior communicating artery, trauma, hydrocephalus, and inflammation.

Hypernatremia can develop following the loss of water via both renal and nonrenal routes. Insensible losses of water may increase in the setting of fever, exercise, heat exposure, severe burns, or mechanical ventilation. Diarrhea is, in turn, the most common gastrointestinal cause of hypernatremia. Notably, osmotic diarrhea and viral gastroenteritis typically generate stools with Na⁺ and K⁺ <100 mM, thus leading to water loss and hypernatremia; in contrast, secretory diarrhea typically results in isotonic stool and thus hypovolemia with or without hypovolemic hyponatremia.

Common causes of renal water loss include osmotic diuresis secondary to hyperglycemia, excess urea, postobstructive diuresis, or mannitol; these disorders share an increase in urinary solute excretion and urinary osmolality (see “Diagnostic Approach,” below). Hypernatremia due to a water diuresis occurs in central or nephrogenic DI (NDI).

NDI is characterized by renal resistance to AVP, which can be partial or complete (see “Diagnostic Approach,” below). Genetic causes include loss-of-function mutations in the X-linked V₂ receptor; mutations in the AVP-responsive aquaporin-2 water channel can cause autosomal recessive and autosomal dominant NDI, whereas recessive deficiency of the aquaporin-1 water channel causes a more modest concentrating defect (Fig. 53-2). Hypercalcemia can also cause polyuria and NDI; calcium signals directly through the calcium-sensing receptor to downregulate Na⁺, K⁺, and Cl⁻ transport by the TALH and water transport in principal cells, thus reducing renal concentrating ability in hypercalcemia. Another common acquired cause of NDI is hypokalemia, which inhibits the renal response to AVP and downregulates aquaporin-2 expression. Several drugs can cause acquired NDI, in particular, lithium, ifosfamide, and several antiviral agents. Lithium causes NDI by multiple mechanisms, including direct inhibition of renal glycogen synthase kinase-3 (GSK3), a kinase thought to be the pharmacologic target of lithium in bipolar disease; GSK3 is required for the response of principal cells to AVP. The entry of lithium through the amiloride-sensitive Na⁺ channel ENaC (Fig. 53-4) is required for the effect of the drug on principal cells, such that combined therapy within lithium and amiloride can mitigate lithium-associated NDI. However, lithium causes chronic tubulointerstitial scarring and chronic kidney disease after prolonged therapy, such that patients may have a persistent NDI long after stopping the drug, with a reduced therapeutic benefit from amiloride.

Finally, gestational DI is a rare complication of late-term pregnancy wherein increased activity of a circulating placental protease with “vasopressinase” activity leads to reduced circulating AVP and polyuria, often accompanied by hypernatremia. DDAVP is an effective therapy for this syndrome, given its resistance to the vasopressinase enzyme.

Clinical Features Hypernatremia increases osmolality of the ECF, generating an osmotic gradient between the ECF and ICF, an efflux of intracellular water, and cellular shrinkage. As in hyponatremia, the symptoms of hypernatremia are predominantly neurologic. Altered mental status is the most frequent manifestation, ranging from mild confusion and lethargy to deep coma. The sudden shrinkage of brain cells in acute hypernatremia may lead to parenchymal or subarachnoid hemorrhages and/or subdural hematomas; however, these vascular complications are primarily encountered in pediatric and neonatal patients. Rarely, osmotic demyelination may occur in acute hypernatremia. Osmotic damage to muscle membranes can also lead to hypernatremic rhabdomyolysis. Brain cells accommodate to a chronic increase in ECF osmolality (>48 h) by activating membrane transporters that mediate influx and intracellular accumulation of organic osmolytes (creatine, betaine, glutamate, myoinositol, and taurine); this results in an increase in ICF water and normalization of brain parenchymal volume. In consequence, patients with *chronic* hypernatremia are less likely to develop severe neurologic compromise. However, the cellular response

to chronic hypernatremia predisposes pediatric patients with hypernatremia, particularly infants, to the development of cerebral edema and seizures during overly rapid hydration (overcorrection of plasma Na^+ concentration by $>10 \text{ mM/d}$). In critically ill adults, however, recent evidence does not indicate that rapid correction of hypernatremia is associated with a higher risk for mortality, seizure, alteration of consciousness, and/or cerebral edema. Given that restricting the rate of correction to $<10 \text{ mM/d}$ has no physiologic sequelae, it seems prudent to restrict correction in adults to this rate; however, should that rate be exceeded, hypernatremia does not need to be reinduced.

Diagnostic Approach The history should focus on the presence or absence of thirst, polyuria, and/or an extrarenal source for water loss, such as diarrhea. The physical examination should include a detailed neurologic exam and an assessment of the ECFV; patients with a particularly large water deficit and/or a combined deficit in electrolytes and water may be hypovolemic, with reduced JVP and orthostasis. Accurate documentation of daily fluid intake and daily urine output is also critical for the diagnosis and management of hypernatremia.

Laboratory investigation should include a measurement of serum and urine osmolality, in addition to urine electrolytes. The appropriate response to hypernatremia and a serum osmolality $>295 \text{ mOsm/kg}$ is an increase in circulating AVP and the excretion of low volumes ($<500 \text{ mL/d}$) of maximally concentrated urine, i.e., urine with osmolality $>800 \text{ mOsm/kg}$; should this be the case, then an extrarenal source of water loss is primarily responsible for the generation of hypernatremia. Many patients with hypernatremia are polyuric; should an osmotic diuresis be responsible, with excessive excretion of $\text{Na}^+ \text{-Cl}^-$, glucose, and/or urea, then daily solute excretion will be $>750\text{-}1000 \text{ mOsm/d}$ ($>15 \text{ mOsm/kg}$ body water per day) (Fig. 53-6). More commonly, patients with hypernatremia and polyuria will have a predominant water diuresis, with excessive excretion of hypotonic, dilute urine.

Adequate differentiation between nephrogenic and central causes of DI requires the measurement of the response in urinary osmolality to DDAVP, combined with measurement of circulating AVP in the setting of hypertonicity. If measurement of serum copeptin is available, an “indirect water deprivation” test can be performed in patients with hypotonic polyuria without hypernatremia; if an infusion of hypertonic saline increases the level of circulating copeptin, a peptide co-secreted with AVP, then the patient suffers from polydipsia rather than central DI. By definition, patients with baseline hypernatremia are hypertonic, with an adequate stimulus for AVP by the posterior pituitary. Therefore, in contrast to polyuric patients with a normal or reduced baseline plasma Na^+ concentration and osmolality, a water deprivation test (Chap. 52) is unnecessary in hypernatremia; indeed, water deprivation is absolutely contraindicated in this setting, given the risk for worsening the hypernatremia. Hypernatremic patients with NDI will have high serum levels of AVP and copeptin. Their low urine osmolality will also fail to respond to DDAVP, increasing by $<50\%$ or $<150 \text{ mOsm/kg}$ from baseline; patients with central DI will respond to DDAVP, with a reduced circulating AVP and copeptin. Patients may exhibit a partial response to DDAVP, with a $>50\%$ rise in urine osmolality that nonetheless fails to reach 800 mOsm/kg ; the level of circulating AVP will help differentiate the underlying cause, i.e., NDI versus central DI. In pregnant patients, AVP assays should be drawn in tubes containing the protease inhibitor 1,10-phenanthroline to prevent in vitro degradation of AVP by placental vasopressinase.

For patients with hypernatremia due to renal loss of water, it is critical to quantify ongoing daily losses using the calculated electrolyte-free water clearance, in addition to calculation of the baseline water deficit (the relevant formulas are discussed in Table 53-3). This requires daily measurement of urine electrolytes, combined with accurate measurement of daily urine volume.

TREATMENT

Hypernatremia

The underlying cause of hypernatremia should be withdrawn or corrected, be it drugs, hyperglycemia, hypercalcemia, hypokalemia, or diarrhea. The approach to the correction of hypernatremia is

outlined in Table 53-3. It is imperative to correct hypernatremia slowly to avoid cerebral edema, typically replacing the calculated free water deficit over 48 h. Notably, the plasma Na^+ concentration should be corrected by no more than 10 mM/d , which may take longer than 48 h in patients with severe hypernatremia ($>160 \text{ mM}$). A rare exception is patients with acute hypernatremia ($<48 \text{ h}$) due to sodium loading, who can safely be corrected rapidly at a rate of 1 mM/h .

Water should ideally be administered by mouth or by nasogastric tube, as the most direct way to provide free water, i.e., water without electrolytes. Alternatively, patients can receive free water in dextrose-containing IV solutions, such as 5% dextrose (D_5W); blood glucose should be monitored in case hyperglycemia occurs. Depending on the history, blood pressure, or clinical volume status, it may be appropriate to initially treat with hypotonic saline solutions (1/4 or 1/2 normal saline); normal saline is usually inappropriate in the absence of very severe hypernatremia, where normal saline is proportionally more hypotonic relative to plasma, or frank hypotension. Calculation of urinary electrolyte-free water clearance (Table 53-3) is required to estimate daily, ongoing loss of free water in patients with NDI or central DI, which should be replenished daily.

Additional therapy may be feasible in specific cases. Patients with central DI should respond to the administration of intravenous, intranasal, or oral DDAVP. Patients with NDI due to lithium may reduce their polyuria with amiloride (2.5–10 mg/d), which decreases entry of lithium into principal cells by inhibiting ENaC (see above); in practice, however, most patients with lithium-associated DI are able to compensate for their polyuria by simply increasing their daily water intake. Thiazides may reduce polyuria due to NDI, ostensibly by inducing hypovolemia and increasing proximal tubular water reabsorption. Occasionally, nonsteroidal anti-inflammatory drugs (NSAIDs) have been used to treat polyuria associated with NDI, reducing the negative effect of intrarenal prostaglandins on urinary concentrating mechanisms; however, this assumes the risks of NSAID-associated gastric and/or renal toxicity. Furthermore, it must be emphasized that thiazides, amiloride, and NSAIDs are only appropriate for *chronic* management of polyuria from NDI and have *no* role in the acute management of associated hypernatremia, where the focus is on replacing free water deficits and ongoing free water loss.

POTASSIUM DISORDERS

Homeostatic mechanisms maintain plasma K^+ concentration between 3.5 and 5.0 mM , despite marked variation in dietary K^+ intake. In a healthy individual at steady state, the entire daily intake of potassium is excreted, ~90% in the urine and 10% in the stool; thus, the kidney plays a dominant role in potassium homeostasis. However, >98% of total-body potassium is intracellular, chiefly in muscle; buffering of extracellular K^+ by this large intracellular pool plays a crucial role in the regulation of plasma K^+ concentration. Changes in the exchange and distribution of intra- and extracellular K^+ can thus lead to marked hypo- or hyperkalemia. A corollary is that massive necrosis and the attendant release of tissue K^+ can cause severe hyperkalemia, particularly in the setting of acute kidney injury and reduced excretion of K^+ .

Changes in whole-body K^+ content are primarily mediated by the kidney, which *reabsorbs* filtered K^+ in hypokalemic, K^+ -deficient states and *secretes* K^+ in hyperkalemic, K^+ -replete states. Although K^+ is transported along the entire nephron, it is the principal cells of the connecting segment (CNT) and cortical CD that play a dominant role in renal K^+ secretion, whereas alpha-intercalated cells of the outer medullary CD function in renal tubular reabsorption of filtered K^+ in K^+ -deficient states. In principal cells, apical Na^+ entry via the amiloride-sensitive ENaC generates a lumen-negative potential difference, which drives passive K^+ exit through apical K^+ channels (Fig. 53-4). Two major K^+ channels mediate distal tubular K^+ secretion: the secretory K^+ channel ROMK (renal outer medullary K^+ channel; also known as Kir1.1 or Kcnj1) and the flow-sensitive “big potassium” (BK) or maxi-K

K^+ channel. ROMK is thought to mediate the bulk of constitutive K^+ secretion, whereas increases in distal flow rate and/or genetic absence of ROMK activate K^+ secretion via the BK channel.

An appreciation of the relationship between ENaC-dependent Na^+ entry and distal K^+ secretion (Fig. 53-4) is required for the bedside interpretation of potassium disorders. For example, decreased distal delivery of Na^+ , as occurs in hypovolemic, prerenal states, tends to blunt the ability to excrete K^+ , leading to hyperkalemia; on the other hand, an *increase* in distal delivery of Na^+ and distal flow rate, as occurs after treatment with thiazide and loop diuretics, can enhance K^+ secretion and lead to hypokalemia. Hyperkalemia is also a predictable consequence of drugs that directly inhibit ENaC, due to the role of this Na^+ channel in generating a lumen-negative potential difference. Aldosterone in turn has a major influence on potassium excretion, increasing the activity of ENaC channels and thus amplifying the driving force for K^+ secretion across the luminal membrane of principal cells. Abnormalities in the renin-angiotensin-aldosterone system can thus cause both hypokalemia and hyperkalemia. Notably, however, potassium excess and potassium restriction have opposing, aldosterone-independent effects on the density and activity of apical K^+ channels in the distal nephron, i.e., factors other than aldosterone modulate the renal capacity to secrete K^+ . In addition, potassium restriction and hypokalemia activate aldosterone-independent distal *reabsorption* of filtered K^+ , activating apical H^+/K^+ -ATPase activity in intercalated cells within the outer medullary CD. Reflective perhaps of this physiology, changes in plasma K^+ concentration are not universal in disorders associated with changes in aldosterone activity.

HYPOKALEMIA

Hypokalemia, defined as a plasma K^+ concentration of <3.5 mM, occurs in up to 20% of hospitalized patients. Hypokalemia is associated with a tenfold increase in in-hospital mortality, due to adverse effects on cardiac rhythm, blood pressure, and cardiovascular morbidity. Mechanistically, hypokalemia can be caused by redistribution of K^+ between tissues and the ECF or by renal and nonrenal loss of K^+ (Table 53-4). Systemic hypomagnesemia can also cause treatment-resistant hypokalemia, due to a combination of reduced cellular uptake of K^+ and exaggerated renal secretion. Spurious hypokalemia or “pseudohypokalemia” can occasionally result from in vitro cellular uptake of K^+ after venipuncture, for example, due to profound leukocytosis in acute leukemia.

Redistribution and Hypokalemia Insulin, β_2 -adrenergic activity, thyroid hormone, and alkalosis promote Na^+/K^+ -ATPase-mediated cellular uptake of K^+ , leading to hypokalemia. Inhibition of the passive efflux of K^+ can also cause hypokalemia, albeit rarely; this typically occurs in the setting of systemic inhibition of K^+ channels by toxic barium ions. Exogenous insulin can cause iatrogenic hypokalemia, particularly during the management of K^+ -deficient states such as diabetic ketoacidosis. Alternatively, the stimulation of endogenous insulin can provoke hypokalemia, hypomagnesemia, and/or hypophosphatemia in malnourished patients given a carbohydrate load. Alterations in the activity of the endogenous sympathetic nervous system can cause hypokalemia in several settings, including alcohol withdrawal, hyperthyroidism, acute myocardial infarction, and severe head injury. β_2 agonists, including both bronchodilators and tocolytics (ritodrine), are powerful activators of cellular K^+ uptake; “hidden” sympathomimetics, such as pseudoephedrine and ephedrine in cough syrup or dieting agents, may also cause unexpected hypokalemia. Finally, xanthine-dependent activation of cAMP-dependent signaling, downstream of the β_2 receptor, can lead to hypokalemia, usually in the setting of overdose (theophylline) or marked overingestion (dietary caffeine).

Redistributive hypokalemia can also occur in the setting of hyperthyroidism, with periodic attacks of hypokalemic paralysis (thyrotoxic periodic paralysis [TPP]). Similar episodes of hypokalemic weakness in the absence of thyroid abnormalities occur in *familial* hypokalemic periodic paralysis, usually caused by missense mutations of voltage sensor domains within the α_1 subunit of L-type calcium channels or the skeletal Na^+ channel; these mutations generate an abnormal gating pore

TABLE 53-4 Causes of Hypokalemia

- I. Decreased intake
 - A. Starvation
 - B. Clay ingestion
- II. Redistribution into cells
 - A. Acid-base
 - 1. Metabolic alkalosis
 - B. Hormonal
 - 1. Insulin
 - 2. Increased β_2 -adrenergic sympathetic activity: post-myocardial infarction, head injury
 - 3. β_2 -Adrenergic agonists—bronchodilators, tocolytics
 - 4. α -Adrenergic antagonists
 - 5. Thyrotoxic periodic paralysis
 - 6. Downstream stimulation of Na^+/K^+ -ATPase: theophylline, caffeine
 - C. Anabolic state
 - 1. Vitamin B_{12} or folic acid administration (red blood cell production)
 - 2. Granulocyte-macrophage colony-stimulating factor (white blood cell production)
 - 3. Total parenteral nutrition
 - D. Other
 - 1. Pseudohypokalemia
 - 2. Hypothermia
 - 3. Familial hypokalemic periodic paralysis
 - 4. Barium toxicity: systemic inhibition of “leak” K^+ channels
- III. Increased loss
 - A. Nonrenal
 - 1. Gastrointestinal loss (diarrhea)
 - 2. Integumentary loss (sweat)
 - B. Renal
 - 1. Increased distal flow and distal Na^+ delivery: diuretics, osmotic diuresis, salt-wasting nephropathies
 - 2. Increased secretion of potassium
 - a. Mineralocorticoid excess: primary hyperaldosteronism (aldosterone-producing adenomas, primary or unilateral adrenal hyperplasia, idiopathic hyperaldosteronism due to bilateral adrenal hyperplasia, and adrenal carcinoma), genetic hyperaldosteronism (familial hyperaldosteronism types I/II/III, congenital adrenal hyperplasias), secondary hyperaldosteronism (malignant hypertension, renin-secreting tumors, renal artery stenosis, hypovolemia), Cushing’s syndrome, Bartter’s syndrome, Gitelman’s syndrome
 - b. Apparent mineralocorticoid excess: genetic deficiency of 11 β -dehydrogenase-2 (syndrome of apparent mineralocorticoid excess), inhibition of 11 β -dehydrogenase-2 (glycyrrhetic acid and/or carbenoxolone; itraconazole and posaconazole; licorice, food products, drugs), Liddle’s syndrome (genetic activation of epithelial Na^+ channels)
 - c. Distal delivery of nonreabsorbed anions: vomiting, nasogastric suction, proximal renal tubular acidosis, diabetic ketoacidosis, glue-sniffing (toluene abuse), penicillin derivatives (penicillin, nafcillin, dicloxacillin, ticarcillin, oxacillin, and carbenicillin)
 - 3. Magnesium deficiency

current activated by hyperpolarization. TPP develops more frequently in patients of Asian or Hispanic origin; this shared predisposition has been linked to genetic variation in Kir2.6, a muscle-specific, thyroid hormone-responsive K^+ channel. Genome-wide association studies have also implicated variation in the *KCNJ2* gene, which encodes a related muscle K^+ channel, Kir 2.1, in predisposition to TPP. Patients with TPP typically present with weakness of the extremities and limb girdles, with paralytic episodes that occur most frequently between 1 and 6 A.M. Signs and symptoms of hyperthyroidism are not invariably present. Hypokalemia is usually profound and almost invariably accompanied by hypophosphatemia and hypomagnesemia. The hypokalemia in TPP is also attributed to both direct and indirect activation of the Na^+/K^+ -ATPase, resulting in increased uptake of K^+ by muscle

and other tissues. Increases in β -adrenergic activity play an important role in that high-dose propranolol (3 mg/kg) rapidly reverses the associated hypokalemia, hypophosphatemia, and paralysis. Outward-directed inward-rectifying K^+ current, mediated by KIR channels (primarily Kir2.1 and Kir2.2 tetramers), is also reduced in skeletal muscles of patients with TPP, providing an additional mechanism for hypokalemia. Together with increased Na^+/K^+ -ATPase activity and increased circulating insulin, this reduced KIR current may trigger a “feedforward” cycle of hypokalemia leading to inactivation of muscle Na^+ channels, paradoxical depolarization, and paralysis.

Nonrenal Loss of Potassium The loss of K^+ in sweat is typically low, except under extremes of physical exertion. Direct gastric losses of K^+ due to vomiting or nasogastric suctioning are also minimal; however, the ensuing hypochloremic alkalosis results in persistent kaliuresis due to secondary hyperaldosteronism and bicarbonaturia, i.e., a renal loss of K^+ . Diarrhea is a globally important cause of hypokalemia, given the worldwide prevalence of infectious diarrheal disease. Noninfectious gastrointestinal processes such as celiac disease, ileostomy, villous adenomas, inflammatory bowel disease, colonic pseudo-obstruction (Ogilvie’s syndrome), VIPomas, and chronic laxative abuse can also cause significant hypokalemia; an exaggerated intestinal secretion of potassium by upregulated colonic BK channels has been directly implicated in the pathogenesis of hypokalemia in many of these disorders.

Renal Loss of Potassium Drugs can increase renal K^+ excretion by a variety of different mechanisms. Diuretics are a particularly common cause, due to associated increases in distal tubular Na^+ delivery and distal tubular flow rate, in addition to secondary hyperaldosteronism. Thiazides have a greater effect on plasma K^+ concentration than loop diuretics, despite their lesser natriuretic effect. The diuretic effect of thiazides is largely due to inhibition of the Na^+Cl^- cotransporter NCC in DCT cells. This leads to a direct increase in the delivery of luminal Na^+ to the principal cells immediately downstream in the CNT and cortical CD, which augments Na^+ entry via ENaC, increases the lumen-negative potential difference, and amplifies K^+ secretion. The higher propensity of thiazides to cause hypokalemia may also be secondary to thiazide-associated hypocaliuria, versus the *hypercalciuria* seen with loop diuretics; the increases in downstream luminal calcium in response to loop diuretics inhibit ENaC in principal cells, thus reducing the lumen-negative potential difference and attenuating distal K^+ excretion. High doses of penicillin-related antibiotics (nafcillin, dicloxacillin, ticarcillin, oxacillin, and carbenicillin) can increase obligatory K^+ excretion by acting as nonreabsorbable anions in the distal nephron. Finally, several renal tubular toxins cause renal K^+ and magnesium wasting, leading to hypokalemia and hypomagnesemia; these drugs include aminoglycosides, amphotericin, foscarnet, cisplatin, and ifosfamide (see also “Magnesium Deficiency and Hypokalemia,” below).

Aldosterone activates the ENaC channel in principal cells via multiple synergistic mechanisms, thus increasing the driving force for K^+ excretion. In consequence, increases in aldosterone bioactivity and/or gains in function of aldosterone-dependent signaling pathways are associated with hypokalemia. Increases in circulating aldosterone (hyperaldosteronism) may be primary or secondary. Increased levels of circulating renin in secondary forms of hyperaldosteronism lead to increased angiotensin II and thus aldosterone; renal artery stenosis is perhaps the most frequent cause (Table 53-4). Primary hyperaldosteronism may be genetic or acquired. Hypertension and hypokalemia, due to increases in circulating 11-deoxycorticosterone, occur in patients with congenital adrenal hyperplasia caused by defects in either steroid 11 β -hydroxylase or steroid 17 α -hydroxylase; deficient 11 β -hydroxylase results in associated virilization and other signs of androgen excess, whereas reduced sex steroids in 17 α -hydroxylase deficiency lead to hypogonadism.

The major forms of *isolated* primary genetic hyperaldosteronism are familial hyperaldosteronism type I (FH-I, also known as glucocorticoid-remediable hyperaldosteronism [GRA]) and familial hyperaldosteronism types II and III (FH-II and FH-III), in which aldosterone production is not repressible by exogenous glucocorticoids.

FH-I is caused by a chimeric gene duplication between the homologous 11 β -hydroxylase (*CYP11B1*) and aldosterone synthase (*CYP11B2*) genes, fusing the adrenocorticotrophic hormone (ACTH)-responsive 11 β -hydroxylase promoter to the coding region of aldosterone synthase; this chimeric gene is under the control of ACTH and thus repressible by glucocorticoids. FH-III is caused by mutations in the *KCNJ5* gene, which encodes the G protein-activated inward rectifier K^+ channel 4 (GIRK4); these mutations lead to the acquisition of sodium permeability in the mutant GIRK4 channels, causing an exaggerated membrane depolarization in adrenal glomerulosa cells and the activation of voltage-gated calcium channels. The resulting calcium influx is sufficient to produce aldosterone secretion and cell proliferation, leading to adrenal adenomas and hyperaldosteronism.

Acquired causes of primary hyperaldosteronism include aldosterone-producing adenomas (APAs), primary or unilateral adrenal hyperplasia (PAH), idiopathic hyperaldosteronism (IHA) due to bilateral adrenal hyperplasia, and adrenal carcinoma; APA and IHA account for close to 60% and 40%, respectively, of diagnosed hyperaldosteronism. Acquired somatic mutations in *KCNJ5* or less frequently in the *ATP1A1* (an Na^+/K^+ ATPase α subunit) and *ATP2B3* (a Ca^{2+} ATPase) genes can be detected in APAs; as in FH-III (see above), the exaggerated depolarization of adrenal glomerulosa cells caused by these mutations is implicated in the excessive adrenal proliferation and the exaggerated release of aldosterone.

Random testing of plasma renin activity (PRA) and aldosterone is a helpful screening tool in hypokalemic and/or hypertensive patients, with an aldosterone:PRA ratio of >50 suggestive of primary hyperaldosteronism. Hypokalemia and multiple antihypertensive drugs may alter the aldosterone:PRA ratio by suppressing aldosterone or increasing PRA, leading to a ratio of <50 in patients who do in fact have primary hyperaldosteronism; therefore, the clinical context should always be considered when interpreting these results.

The glucocorticoid cortisol has equal affinity for the MR to that of aldosterone, with resultant “mineralocorticoid-like” activity. However, cells in the aldosterone-sensitive distal nephron are protected from this “illicit” activation by the enzyme 11 β -hydroxysteroid dehydrogenase-2 (11 β HSD-2), which converts cortisol to cortisone; cortisone has minimal affinity for the MR. Recessive loss-of-function mutations in the 11 β HSD-2 gene are thus associated with cortisol-dependent activation of the MR and the syndrome of apparent mineralocorticoid excess (SAME), encompassing hypertension, hypokalemia, hypercalciuria, and metabolic alkalosis, with suppressed PRA and suppressed aldosterone. A similar syndrome is caused by biochemical inhibition of 11 β HSD-2 by glycyrrhetic acid/glycyrrhizic acid and/or carbenoxolone. Glycyrrhetic acid is a natural sweetener found in licorice root, typically encountered in licorice and its many guises or as a flavoring agent in tobacco and food products. More recently, the antifungals itraconazole and posaconazole have been shown to inhibit 11 β HSD-2, leading to hypertension and hypokalemia.

Finally, hypokalemia may also occur with systemic increases in glucocorticoids. In Cushing’s syndrome caused by increases in pituitary ACTH (Chap. 386), the incidence of hypokalemia is only 10%, whereas it is 60–100% in patients with ectopic secretion of ACTH, despite a similar incidence of hypertension. Indirect evidence suggests that the activity of renal 11 β HSD-2 is reduced in patients with ectopic ACTH compared with Cushing’s syndrome, resulting in SAME.

Finally, defects in multiple renal tubular transport pathways are associated with hypokalemia. For example, loss-of-function mutations in subunits of the acidifying H^+ -ATPase in alpha-intercalated cells cause hypokalemic distal renal tubular acidosis, as do many acquired disorders of the distal nephron. Liddle’s syndrome is caused by autosomal dominant gain-in-function mutations of ENaC subunits. Disease-associated mutations either activate the channel directly or abrogate aldosterone-inhibited retrieval of ENaC subunits from the plasma membrane; the end result is increased expression of activated ENaC channels at the plasma membrane of principal cells. Patients with Liddle’s syndrome classically manifest severe hypertension with hypokalemia, unresponsive to spironolactone yet sensitive to amiloride. Hypertension and hypokalemia are, however, variable aspects

of the Liddle's phenotype; more consistent features include a blunted aldosterone response to ACTH and reduced urinary aldosterone excretion.

Loss of the transport functions of the TALH and DCT nephron segments causes hereditary hypokalemic alkalosis and Bartter's syndrome (BS) and Gitelman's syndrome (GS), respectively. Patients with classic BS typically suffer from polyuria and polydipsia, due to the reduction in renal concentrating ability. They may have an increase in urinary calcium excretion, and 20% are hypomagnesemic. Other features include marked activation of the renin-angiotensin-aldosterone axis. Patients with antenatal BS suffer from a severe systemic disorder characterized by marked electrolyte wasting, polyhydramnios, and hypercalciuria with nephrocalcinosis; renal prostaglandin synthesis and excretion are significantly increased, accounting for much of the systemic symptoms. There are five disease genes for BS, all of them functioning in some aspect of regulated Na^+ , K^+ , and Cl^- transport by the TALH. In contrast, GS is genetically homogeneous, caused almost exclusively by loss-of-function mutations in the thiazide-sensitive Na^+/Cl^- cotransporter of the DCT. Patients with GS are uniformly hypomagnesemic and exhibit marked hypocalciuria, rather than the hypercalciuria typically seen in BS; urinary calcium excretion is thus a critical diagnostic test in GS. GS is a milder phenotype than BS; however, patients with GS may suffer from chondrocalcinosis, an abnormal deposition of calcium pyrophosphate dihydrate (CPPD) in joint cartilage (Chap. 315).

Magnesium Deficiency and Hypokalemia Magnesium depletion has inhibitory effects on muscle Na^+/K^+ -ATPase activity, reducing influx into muscle cells and causing a secondary kaliuresis. In addition, magnesium depletion causes exaggerated K^+ secretion by the distal nephron; this effect is attributed to a reduction in the magnesium-dependent, intracellular block of K^+ efflux through the secretory K^+ channel of principal cells (ROMK; Fig. 53-4). In consequence, hypomagnesemic patients are clinically refractory to K^+ replacement in the absence of Mg^{2+} repletion. Notably, magnesium deficiency is also a common concomitant of hypokalemia because many disorders of the distal nephron may cause both potassium and magnesium wasting (Chap. 315).

Clinical Features Hypokalemia has prominent effects on cardiac, skeletal, and intestinal muscle cells. In particular, hypokalemia is a major risk factor for both ventricular and atrial arrhythmias. Hypokalemia predisposes to digoxin toxicity by a number of mechanisms, including reduced competition between K^+ and digoxin for shared binding sites on cardiac Na^+/K^+ -ATPase subunits. Electrocardiographic changes in hypokalemia include broad flat T waves, ST depression, and QT prolongation; these are most marked when serum K^+ is <2.7 mmol/L. Hypokalemia can thus be an important precipitant of arrhythmia in patients with additional genetic or acquired causes of QT prolongation. Hypokalemia also results in hyperpolarization of skeletal muscle, thus impairing the capacity to depolarize and contract; weakness and even paralysis may ensue. It also causes a skeletal myopathy and predisposes to rhabdomyolysis. Finally, the paralytic effects of hypokalemia on intestinal smooth muscle may cause intestinal ileus.

The functional effects of hypokalemia on the kidney can include Na^+/Cl^- and HCO_3^- retention, polyuria, phosphaturia, hypocitraturia, and an activation of renal ammoniogenesis. Bicarbonate retention and other acid-base effects of hypokalemia can contribute to the generation of metabolic alkalosis. Hypokalemic polyuria is due to a combination of central polydipsia and an AVP-resistant renal concentrating defect. Structural changes in the kidney due to hypokalemia include a relatively specific vacuolizing injury to proximal tubular cells, interstitial nephritis, and renal cysts. Hypokalemia also predisposes to acute kidney injury and can lead to end-stage renal disease (ESRD) in patients with long-standing hypokalemia due to eating disorders and/or laxative abuse.

Hypokalemia and/or reduced dietary K^+ are implicated in the pathophysiology and progression of hypertension, heart failure, vascular disease, and stroke. For example, short-term K^+ restriction in healthy humans and patients with essential hypertension induces Na^+/Cl^- retention and hypertension. Correction of hypokalemia is particularly

important in hypertensive patients treated with diuretics, in whom blood pressure improves with potassium supplementation and the establishment of normokalemia.

Diagnostic Approach The cause of hypokalemia is usually evident from history, physical examination, and/or basic laboratory tests. The history should focus on medications (e.g., laxatives, diuretics, antibiotics), diet and dietary habits (e.g., licorice), and/or symptoms that suggest a particular cause (e.g., periodic weakness, diarrhea). The physical examination should pay particular attention to blood pressure, volume status, and signs suggestive of specific hypokalemic disorders, e.g., hyperthyroidism and Cushing's syndrome. Initial laboratory evaluation should include electrolytes, BUN, creatinine, serum osmolality, Mg^{2+} , Ca^{2+} , a complete blood count, and urinary pH, osmolality, creatinine, and electrolytes (Fig. 53-7). The presence of a non-anion gap acidosis suggests a distal, hypokalemic renal tubular acidosis or diarrhea; calculation of the urinary anion gap can help differentiate these two diagnoses. Renal K^+ excretion can be assessed with a 24-h urine collection; a 24-h K^+ excretion of <15 mmol is indicative of an extrarenal cause of hypokalemia (Fig. 53-7). If only a random, spot urine sample is available, serum and urine osmolality can be used to calculate the transtubular K^+ gradient (TTKG), which should be <3 in the presence of hypokalemia (see also "Hyperkalemia"). Alternatively, a urinary K^+ -to-creatinine ratio of >13 mmol/g creatinine (>1.5 mmol/mmol creatinine) is compatible with excessive renal K^+ excretion. Urine Cl^- is usually decreased in patients with hypokalemia from a nonreabsorbable anion, such as antibiotics or HCO_3^- . The most common causes of chronic hypokalemic alkalosis are surreptitious vomiting, diuretic abuse, and GS; these can be distinguished by the pattern of urinary electrolytes. Hypokalemic patients with vomiting due to bulimia will thus typically have a urinary Cl^- <10 mmol/L; urine Na^+ , K^+ , and Cl^- are persistently elevated in GS, due to loss of function in the thiazide-sensitive Na^+/Cl^- cotransporter, but less elevated in diuretic abuse and with greater variability. Urine diuretic screens for loop diuretics and thiazides may be necessary to further exclude diuretic abuse.

Other tests, such as urinary Ca^{2+} , thyroid function tests, and/or PRA and aldosterone levels, may also be appropriate in specific cases. A plasma aldosterone:PRA ratio of >50, due to suppression of circulating renin and an elevation of circulating aldosterone, is suggestive of hyperaldosteronism. Patients with hyperaldosteronism or apparent mineralocorticoid excess may require further testing, for example, adrenal vein sampling (Chap. 386) or the clinically available testing for specific genetic causes (e.g., FH-I, SAME, Liddle's syndrome). Patients with primary aldosteronism should thus be tested for the chimeric FH-I/GRA gene (see above) if they are younger than 20 years of age or have a family history of primary aldosteronism or stroke at a young age (<40 years). Preliminary differentiation of Liddle's syndrome due to mutant ENaC channels from SAME due to mutant 11 β HSD-2 (see above), both of which cause hypokalemia and hypertension with aldosterone suppression, can be made on a clinical basis and then confirmed by genetic analysis; patients with Liddle's syndrome should respond to amiloride (ENaC inhibition) but not spironolactone, whereas patients with SAME will respond to spironolactone.

TREATMENT

Hypokalemia

The goals of therapy in hypokalemia are to prevent life-threatening and/or serious chronic consequences, to replace the associated K^+ deficit, and to correct the underlying cause and/or mitigate future hypokalemia. The urgency of therapy depends on the severity of hypokalemia, associated clinical factors (e.g., cardiac disease, digoxin therapy), and the rate of decline in serum K^+ . Patients with a prolonged QT interval and/or other risk factors for arrhythmia should be monitored by continuous cardiac telemetry during repletion. Urgent but cautious K^+ replacement should be considered in patients with severe redistributive hypokalemia (plasma K^+ concentration <2.5 mM) and/or when serious complications ensue; however, this approach has a risk of rebound hyperkalemia following

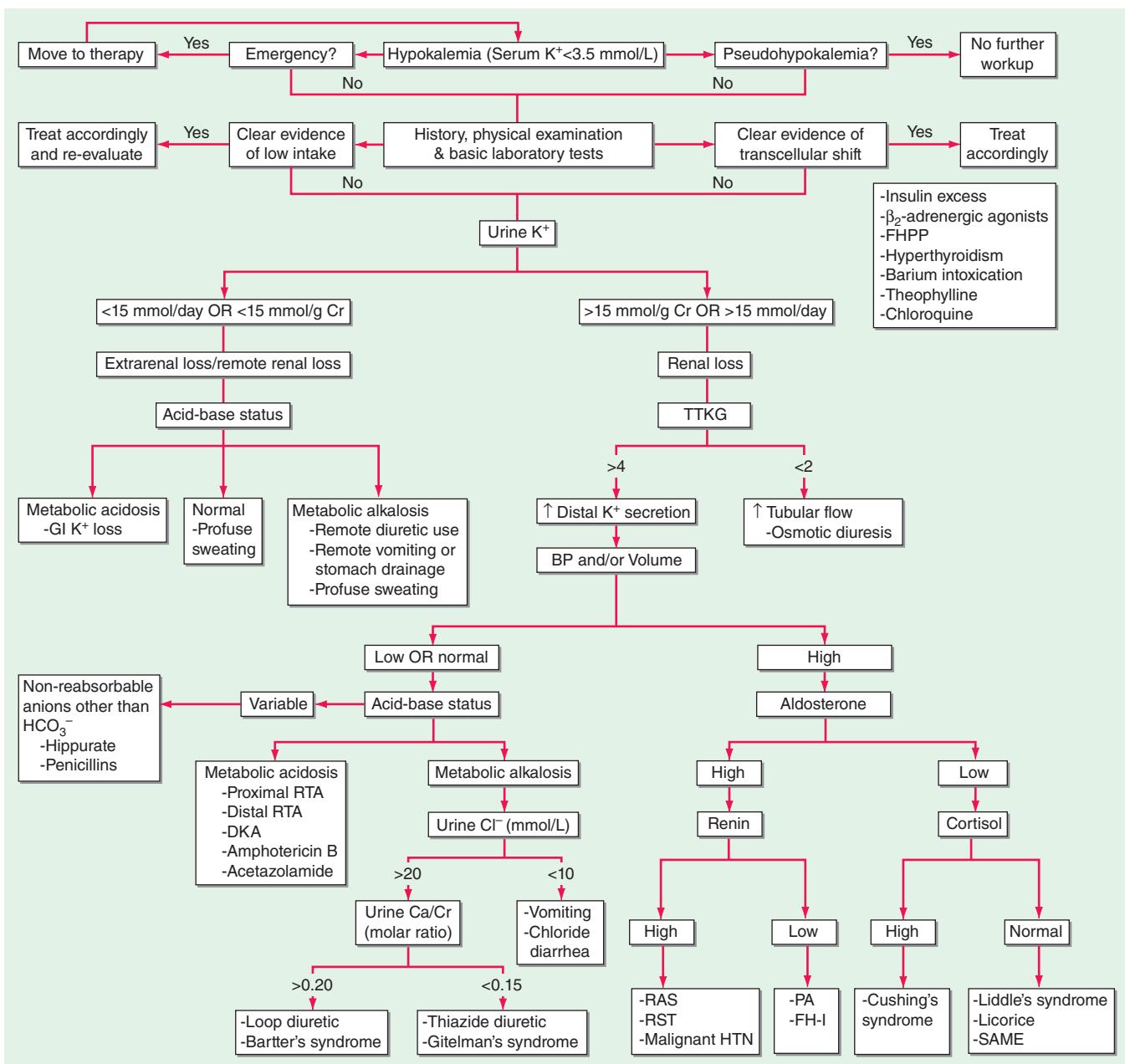


FIGURE 53-7 The diagnostic approach to hypokalemia. See text for details. AME, apparent mineralocorticoid excess; BP, blood pressure; CCD, cortical collecting duct; DKA, diabetic ketoacidosis; FH-I, familial hyperaldosteronism type I; FHPP, familial hypokalemic periodic paralysis; GI, gastrointestinal; GRA, glucocorticoid remediable aldosteronism; HTN, hypertension; PA, primary aldosteronism; RAS, renal artery stenosis; RST, renin-secreting tumor; RTA, renal tubular acidosis; SAME, syndrome of apparent mineralocorticoid excess; TTKG, transtubular potassium gradient. (Reproduced with permission from DB Mount, K Zandi-Nejad: Disorders of potassium balance, in BM Brenner [ed], Brenner and Rector's The Kidney, 8th ed, Philadelphia, W.B. Saunders & Company, 2008.)

acute resolution of the underlying cause. When excessive activity of the sympathetic nervous system is thought to play a dominant role in redistributive hypokalemia, as in TPP, theophylline overdose, and acute head injury, high-dose propranolol (3 mg/kg) should be considered; this nonspecific β-adrenergic blocker will correct hypokalemia without the risk of rebound hyperkalemia.

Oral replacement with K⁺-Cl⁻ is the mainstay of therapy in hypokalemia. Potassium phosphate, oral or IV, may be appropriate in patients with combined hypokalemia and hypophosphatemia. Potassium bicarbonate or potassium citrate should be considered in patients with concomitant metabolic acidosis. Notably, hypomagnesemic patients are refractory to K⁺ replacement alone, such that concomitant Mg²⁺ deficiency should *always* be corrected with oral or intravenous repletion. The deficit of K⁺ and the rate of correction should be estimated as accurately as possible; renal function, medications, and comorbid conditions such as diabetes should

also be considered, so as to gauge the risk of overcorrection. In the absence of abnormal K⁺ redistribution, the total deficit correlates with serum K⁺, such that serum K⁺ drops by ~0.27 mM for every 100-mmol reduction in total-body stores; loss of 400–800 mmol of total-body K⁺ results in a reduction in serum K⁺ by ~2.0 mM. Notably, given the delay in redistributing potassium into intracellular compartments, this deficit must be replaced gradually over 24–48 h, with frequent monitoring of plasma K⁺ concentration to avoid transient overrepletion and transient hyperkalemia.

The use of intravenous administration should be limited to patients unable to use the enteral route or in the setting of severe complications (e.g., paralysis, arrhythmia). Intravenous K⁺-Cl⁻ should always be administered in saline solutions, rather than dextrose, because the dextrose-induced increase in insulin can acutely exacerbate hypokalemia. The peripheral intravenous dose is usually 20–40 mmol of K⁺-Cl⁻ per liter; higher concentrations

can cause localized pain from chemical phlebitis, irritation, and sclerosis. If hypokalemia is severe (<2.5 mmol/L) and/or critically symptomatic, intravenous K⁺-Cl⁻ can be administered through a central vein with cardiac monitoring in an intensive care setting, at rates of 10–20 mmol/h; higher rates should be reserved for acutely life-threatening complications. The absolute amount of administered K⁺ should be restricted (e.g., 20 mmol in 100 mL of saline solution) to prevent inadvertent infusion of a large dose.

Strategies to minimize K⁺ losses should also be considered. These measures may include minimizing the dose of non-K⁺-sparing diuretics, restricting Na⁺ intake, and using clinically appropriate combinations of non-K⁺-sparing and K⁺-sparing medications (e.g., loop diuretics with ACE inhibitors).

HYPERKALEMIA

Hyperkalemia is defined as a plasma potassium level of 5.5 mM, occurring in up to 10% of hospitalized patients; severe hyperkalemia (>6.0 mM) occurs in ~1%, with a significantly increased risk of mortality. Although redistribution and reduced tissue uptake can acutely cause hyperkalemia, a decrease in renal K⁺ excretion is the most frequent underlying cause (Table 53-5). Excessive intake of K⁺ is a rare cause, given the adaptive capacity to increase renal secretion; however, dietary intake can have a major effect in susceptible patients, e.g., diabetics with hyporeninemic hypoaldosteronism and chronic kidney disease. Drugs that impact on the renin-angiotensin-aldosterone axis are also a major cause of hyperkalemia.

Pseudohyperkalemia Hyperkalemia should be distinguished from factitious hyperkalemia or “pseudohyperkalemia,” an artifactual increase in serum K⁺ due to the release of K⁺ during or after venipuncture. Pseudohyperkalemia can occur in the setting of excessive muscle activity during venipuncture (e.g., fist clenching), a marked increase in cellular elements (thrombocytosis, leukocytosis, and/or erythrocytosis) with in vitro efflux of K⁺, and acute anxiety during venipuncture with respiratory alkalosis and redistributive hyperkalemia. Cooling of blood following venipuncture is another cause, due to reduced cellular uptake; the converse is the increased uptake of K⁺ by cells at high ambient temperatures, leading to normal values for hyperkalemic patients and/or to spurious hypokalemia in normokalemic patients. Finally, there are multiple genetic subtypes of hereditary pseudohyperkalemia, caused by increases in the passive K⁺ permeability of erythrocytes. For example, causative mutations have been described in the red cell anion exchanger (AE1, encoded by the *SLC4A1* gene), leading to reduced red cell anion transport, hemolytic anemia, the acquisition of a novel AE1-mediated K⁺ leak, and pseudohyperkalemia.

Redistribution and Hyperkalemia Several different mechanisms can induce an efflux of intracellular K⁺ and hyperkalemia. Acidemia is associated with cellular uptake of H⁺ and an associated efflux of K⁺; it is thought that this effective K⁺-H⁺ exchange serves to help maintain extracellular pH. Notably, this effect of acidosis is limited to non-anion gap causes of metabolic acidosis and, to a lesser extent, respiratory causes of acidosis; hyperkalemia due to an acidosis-induced shift of potassium from the cells into the ECF does *not* occur in the anion gap acidoses lactic acidosis and ketoacidosis. Hyperkalemia due to hypertonic mannitol, hypertonic saline, and intravenous immune globulin is generally attributed to a “solvent drag” effect, as water moves out of cells along the osmotic gradient. Diabetics are also prone to osmotic hyperkalemia in response to intravenous hypertonic glucose, when given without adequate insulin. Cationic amino acids, specifically lysine, arginine, and the structurally related drug epsilon-aminocaproic acid, cause efflux of K⁺ and hyperkalemia, through an effective cation-K⁺ exchange of unknown identity and mechanism. Digoxin inhibits Na⁺/K⁺-ATPase and impairs the uptake of K⁺ by skeletal muscle, such that digoxin overdose predictably results in hyperkalemia. Structurally related glycosides are found in specific plants (e.g., yellow oleander, foxglove) and in the cane toad, *Bufo marinus* (bufadienolide); ingestion of these substances and extracts thereof can also

TABLE 53-5 Causes of Hyperkalemia

- I. Pseudohyperkalemia
 - A. Cellular efflux; thrombocytosis, erythrocytosis, leukocytosis, in vitro hemolysis
 - B. Hereditary defects in red cell membrane transport
- II. Intra- to extracellular shift
 - A. Acidosis
 - B. Hyperosmolality; radiocontrast, hypertonic dextrose, mannitol
 - C. β_2 -Adrenergic antagonists (noncardioselective agents)
 - D. Digoxin and related glycosides (yellow oleander, foxglove, bufadienolide)
 - E. Hyperkalemic periodic paralysis
 - F. Lysine, arginine, and ϵ -aminocaproic acid (structurally similar, positively charged)
 - G. Succinylcholine; thermal trauma, neuromuscular injury, disuse atrophy, mucositis, or prolonged immobilization
 - H. Rapid tumor lysis
- III. Inadequate excretion
 - A. Inhibition of the renin-angiotensin-aldosterone axis; ↑ risk of hyperkalemia when used in combination
 - 1. Angiotensin-converting enzyme (ACE) inhibitors
 - 2. Renin inhibitors; aliskiren (in combination with ACE inhibitors or angiotensin receptor blockers [ARBs])
 - 3. ARBs
 - 4. Blockade of the mineralocorticoid receptor: spironolactone, eplerenone, drospirenone
 - 5. Blockade of the epithelial sodium channel (ENaC): amiloride, triamterene, trimethoprim, pentamidine, nafamostat
 - B. Decreased distal delivery
 - 1. Congestive heart failure
 - 2. Volume depletion
 - C. Hyporeninemic hypoaldosteronism
 - 1. Tubulointerstitial diseases: systemic lupus erythematosus (SLE), sickle cell anemia, obstructive uropathy
 - 2. Diabetes, diabetic nephropathy
 - 3. Drugs: nonsteroidal anti-inflammatory drugs (NSAIDs), cyclooxygenase 2 (COX2) inhibitors, β blockers, cyclosporine, tacrolimus
 - 4. Chronic kidney disease, advanced age
 - 5. Pseudohypoaldosteronism type II: defects in WNK1 or WNK4 kinases, Kelch-like 3 (KLHL3), or Culin 3 (CUL3)
 - D. Renal resistance to mineralocorticoid
 - 1. Tubulointerstitial diseases: SLE, amyloidosis, sickle cell anemia, obstructive uropathy, post-acute tubular necrosis
 - 2. Hereditary: pseudohypoaldosteronism type I; defects in the mineralocorticoid receptor or the epithelial sodium channel (ENaC)
 - E. Advanced renal insufficiency
 - 1. Chronic kidney disease
 - 2. End-stage renal disease
 - 3. Acute oliguric kidney injury
 - F. Primary adrenal insufficiency
 - 1. Autoimmune: Addison's disease, polyglandular endocrinopathy
 - 2. Infectious: HIV, cytomegalovirus, tuberculosis, disseminated fungal infection
 - 3. Infiltrative: amyloidosis, malignancy, metastatic cancer
 - 4. Drug-associated: heparin, low-molecular-weight heparin
 - 5. Hereditary: adrenal hypoplasia congenita, congenital lipoid adrenal hyperplasia, aldosterone synthase deficiency
 - 6. Adrenal hemorrhage or infarction, including in antiphospholipid syndrome

cause hyperkalemia. Finally, fluoride ions also inhibit Na⁺/K⁺-ATPase, such that fluoride poisoning is typically associated with hyperkalemia.

Succinylcholine depolarizes muscle cells, causing an efflux of K⁺ through acetylcholine receptors (AChRs). The use of this agent is contraindicated in patients who have sustained thermal trauma, neuromuscular injury, disuse atrophy, mucositis, or prolonged immobilization.

These disorders share a marked increase and redistribution of AChRs at the plasma membrane of muscle cells; depolarization of these upregulated AChRs by succinylcholine leads to an exaggerated efflux of K⁺ through the receptor-associated cation channels, resulting in acute hyperkalemia.

Hyperkalemia Caused by Excess Intake or Tissue Necrosis

Increased intake of even small amounts of K⁺ may provoke severe hyperkalemia in patients with predisposing factors; hence, an assessment of dietary intake is crucial. Foods rich in potassium include tomatoes, bananas, and citrus fruits; occult sources of K⁺, particularly K⁺-containing salt substitutes, may also contribute significantly. Iatrogenic causes include simple overreplacement with K⁺-Cl⁻ or the administration of a potassium-containing medication (e.g., K⁺-penicillin) to a susceptible patient. Red cell transfusion is a well-described cause of hyperkalemia, typically in the setting of massive transfusions. Finally, severe tissue necrosis, as in acute tumor lysis syndrome and rhabdomyolysis, will predictably cause hyperkalemia from the release of intracellular K⁺.

Hypoaldosteronism and Hyperkalemia Aldosterone release from the adrenal gland may be reduced by hyporeninemic hypoaldosteronism, medications, primary hypoaldosteronism, or isolated deficiency of ACTH (secondary hypoaldosteronism). Primary hypoaldosteronism may be genetic or acquired (Chap. 386) but is commonly caused by autoimmunity, either in Addison's disease or in the context of a polyglandular endocrinopathy. HIV has surpassed tuberculosis as the most important infectious cause of adrenal insufficiency. The adrenal involvement in HIV disease is usually subclinical; however, adrenal insufficiency may be precipitated by stress, drugs such as ketoconazole that inhibit steroidogenesis, or the acute withdrawal of steroid agents such as megestrol. Among medications associated with hyperkalemia, heparin preparations can cause selective inhibition of aldosterone synthesis by zona glomerulosa cells, leading to hyporeninemic hypoaldosteronism.

Hyporeninemic hypoaldosteronism is a very common predisposing factor in several overlapping subsets of hyperkalemic patients: diabetics, the elderly, and patients with renal insufficiency. Classically, patients should have suppressed PRA and aldosterone; ~50% have an associated acidosis, with a reduced renal excretion of NH₄⁺, a positive urinary anion gap, and urine pH <5.5. Most patients are volume expanded, with secondary increases in circulating atrial natriuretic peptide (ANP) that inhibit both renal renin release and adrenal aldosterone release.

Renal Disease and Hyperkalemia Chronic kidney disease and end-stage kidney disease are very common causes of hyperkalemia, due to the associated deficit or absence of functioning nephrons. Hyperkalemia is more common in oliguric acute kidney injury; distal tubular flow rate and Na⁺ delivery are less limiting factors in nonoliguric patients. Hyperkalemia out of proportion to GFR can also be seen in the context of tubulointerstitial disease that affects the distal nephron, such as amyloidosis, sickle cell anemia, interstitial nephritis, and obstructive uropathy.

Hereditary renal causes of hyperkalemia have overlapping clinical features with hypoaldosteronism, hence the diagnostic label *pseudo-hypoaldosteronism* (PHA). PHA type I (PHA-I) has both an autosomal recessive and an autosomal dominant form. The autosomal dominant form is due to loss-of-function mutations in the MLR; the recessive form is caused by various combinations of mutations in the three subunits of ENaC, resulting in impaired Na⁺ channel activity in principal cells and other tissues. Patients with recessive PHA-I suffer from lifelong salt wasting, hypotension, and hyperkalemia, whereas the phenotype of autosomal dominant PHA-I due to MLR dysfunction improves in adulthood. PHA type II (PHA-II; also known as *hereditary hypertension with hyperkalemia*) is in every respect the mirror image of GS caused by loss of function in NCC, the thiazide-sensitive Na⁺-Cl⁻ cotransporter (see above); the clinical phenotype includes hypertension, hyperkalemia, hyperchloremic metabolic acidosis, suppressed PRA and aldosterone, hypercalciuria, and reduced bone density.

PHA-II thus behaves like a gain of function in NCC, and treatment with thiazides results in resolution of the entire clinical phenotype. However, the NCC gene is not directly involved in PHA-II, which is caused by mutations in the WNK1 and WNK4 serine-threonine kinases or the upstream Kelch-like 3 (KLHL3) and Cullin 3 (CUL3) proteins, two components of an E3 ubiquitin ligase complex that regulates these kinases; these proteins collectively regulate NCC activity, with PHA-II-associated activation of the transporter.

Medication-Associated Hyperkalemia Most medications associated with hyperkalemia cause inhibition of some component of the renin-angiotensin-aldosterone axis. ACE inhibitors, angiotensin receptor blockers, renin inhibitors, and MRAs are predictable and common causes of hyperkalemia, particularly when prescribed in combination. The oral contraceptive agent Yasmin-28 contains the progestin drospirenone, which inhibits the MLR and can cause hyperkalemia in susceptible patients. Cyclosporine, tacrolimus, NSAIDs, and cyclooxygenase 2 (COX2) inhibitors cause hyperkalemia by multiple mechanisms, but share the ability to cause hyporeninemic hypoaldosteronism. Notably, most drugs that affect the renin-angiotensin-aldosterone axis also block the local adrenal response to hyperkalemia, thus attenuating the *direct* stimulation of aldosterone release by increased plasma K⁺ concentration.

Inhibition of apical ENaC activity in the distal nephron by amiloride and other K⁺-sparing diuretics results in hyperkalemia, often with a voltage-dependent hyperchloremic acidosis and/or hypovolemic hyponatremia. Amiloride is structurally similar to the antibiotics TMP and pentamidine, which also block ENaC; risk factors for TMP-associated hyperkalemia include the administered dose, renal insufficiency, and hyporeninemic hypoaldosteronism. Indirect inhibition of ENaC at the plasma membrane is also a cause of drug-associated hyperkalemia; nafamostat, a protease inhibitor used in some countries for anticoagulation and for the management of pancreatitis, inhibits aldosterone-induced renal proteases that activate ENaC by proteolytic cleavage.

Clinical Features Hyperkalemia is a medical emergency due to its effects on the heart. Cardiac arrhythmias associated with hyperkalemia include sinus bradycardia, sinus arrest, slow idioventricular rhythms, ventricular tachycardia, ventricular fibrillation, and asystole. Mild increases in extracellular K⁺ affect the repolarization phase of the cardiac action potential, resulting in changes in T-wave morphology; further increase in plasma K⁺ concentration depresses intracardiac conduction, with progressive prolongation of the PR and QRS intervals. Severe hyperkalemia results in loss of the P wave and a progressive widening of the QRS complex; development of a sine-wave sinoventricular rhythm suggests impending ventricular fibrillation or asystole. Hyperkalemia can also cause a type I Brugada pattern in the electrocardiogram (ECG), with a pseudo-right bundle branch block and persistent coved ST-segment elevation in at least two precordial leads. This hyperkalemic Brugada's sign occurs in critically ill patients with severe hyperkalemia and can be differentiated from genetic Brugada's syndrome by an absence of P waves, marked QRS widening, and an abnormal QRS axis. Classically, the ECG manifestations in hyperkalemia progress from tall peaked T waves (5.5–6.5 mM), to a loss of P waves (6.5–7.5 mM), to a widened QRS complex (7.0–8.0 mM), and, ultimately, a to a sine wave pattern (>8.0 mM). However, these changes are notoriously insensitive, particularly in patients with chronic kidney disease or ESRD.

Hyperkalemia from a variety of causes can also present with ascending paralysis, denoted *secondary hyperkalemic paralysis* to differentiate it from familial hyperkalemic periodic paralysis (HYPP). The presentation may include diaphragmatic paralysis and respiratory failure. Patients with familial HYPP develop myopathic weakness during hyperkalemia induced by increased K⁺ intake or rest after heavy exercise. Depolarization of skeletal muscle by hyperkalemia unmasks an inactivation defect in skeletal Na⁺ channel; autosomal dominant mutations in the SCN4A gene encoding this channel are the predominant cause.

Within the kidney, hyperkalemia has negative effects on the ability to excrete an acid load, such that hyperkalemia per se can contribute to

metabolic acidosis. This defect appears to be due in part to competition between K^+ and NH_4^+ for reabsorption by the TALH and subsequent countercurrent multiplication, ultimately reducing the medullary gradient for NH_3/NH_4 excretion by the distal nephron. Regardless of the underlying mechanism, restoration of normokalemia can, in many instances, correct hyperkalemic metabolic acidosis.

Diagnostic Approach The first priority in the management of hyperkalemia is to assess the need for emergency treatment, followed by a comprehensive workup to determine the cause (Fig. 53-8). History and physical examination should focus on medications, diet and dietary supplements, risk factors for kidney failure, reduction in urine output, blood pressure, and volume status. Initial laboratory tests should include electrolytes, BUN, creatinine, serum osmolality, Mg^{2+} and Ca^{2+} , a complete blood count, and urinary pH, osmolality, creatinine, and electrolytes. A urine Na^+ concentration of <20 mM indicates

that distal Na^+ delivery is a limiting factor in K^+ excretion; volume repletion with 0.9% saline or treatment with furosemide may be effective in reducing plasma K^+ concentration. Serum and urine osmolality are required for calculation of the transtubular K^+ gradient (TTKG) (Fig. 53-8). The expected values of the TTKG are largely based on historical data, and are <3 in the presence of hypokalemia and $>7-8$ in the presence of hyperkalemia. Notably, some authors have opined that the TTKG does not consider the effects of distal tubular urea reabsorption on potassium excretion, concluding that the TTKG is, thus, an unreliable test in the assessment of hyperkalemia. These criticisms are theoretical and not supported by animal experiments; the TTKG remains a helpful bedside test of urinary potassium excretion in hyperkalemia.

$$\text{TTKG} = \frac{[K^+]_{\text{urine}} \times \text{Osm}_{\text{serum}}}{[K^+]_{\text{serum}} \times \text{Osm}_{\text{urine}}}$$

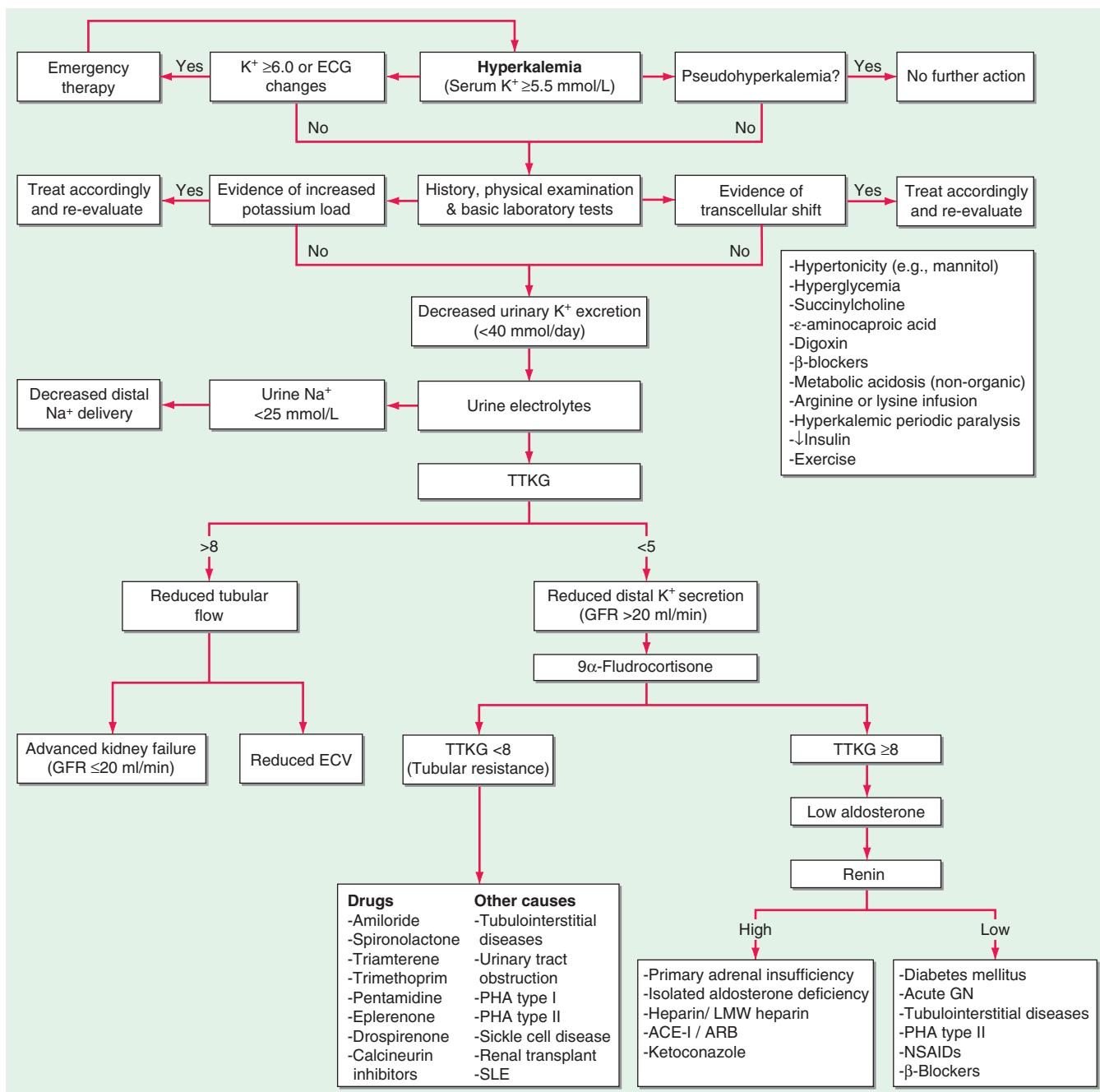


FIGURE 53-8 The diagnostic approach to hyperkalemia. See text for details. ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker; CCD, cortical collecting duct; ECG, electrocardiogram; ECV, effective circulatory volume; GFR, glomerular filtration rate; GN, glomerulonephritis; HIV, human immunodeficiency virus; LMW heparin, low-molecular-weight heparin; NSAIDs, nonsteroidal anti-inflammatory drugs; PHA, pseudohypoaldosteronism; SLE, systemic lupus erythematosus; TTKG, transtubular potassium gradient. (Reproduced with permission from DB Mount, K Zandi-Nejad: Disorders of potassium balance, in BM Brenner [ed], *Brenner and Rector's The Kidney*, 8th ed, Philadelphia, W.B. Saunders & Company, 2008.)

TREATMENT

Hyperkalemia

ECG manifestations of hyperkalemia should be considered a medical emergency and treated urgently. However, patients with significant hyperkalemia (plasma K⁺ concentration ≥ 6.5 mM) in the absence of ECG changes should also be aggressively managed, given the limitations of ECG changes as a predictor of cardiac toxicity. Urgent management of hyperkalemia includes admission to the hospital, continuous cardiac monitoring, and immediate treatment. The treatment of hyperkalemia is divided into three stages:

- 1. Immediate antagonism of the cardiac effects of hyperkalemia.** Intravenous calcium serves to protect the heart, whereas other measures are taken to correct hyperkalemia. Calcium raises the action potential threshold and reduces excitability, without changing the resting membrane potential. By restoring the difference between resting and threshold potentials, calcium reverses the depolarization blockade due to hyperkalemia. The recommended dose is 10 mL of 10% calcium gluconate (3–4 mL of calcium chloride), infused intravenously over 2–3 min with cardiac monitoring. The effect of the infusion starts in 1–3 min and lasts 30–60 min; the dose should be repeated if there is no change in ECG findings or if they recur after initial improvement. Hypercalcemia potentiates the cardiac toxicity of digoxin; hence, intravenous calcium should be used with extreme caution in patients taking this medication; if judged necessary, 10 mL of 10% calcium gluconate can be added to 100 mL of 5% dextrose in water and infused over 20–30 min to avoid acute hypercalcemia.
- 2. Rapid reduction in plasma K⁺ concentration by redistribution into cells.** Insulin lowers plasma K⁺ concentration by shifting K⁺ into cells. The recommended dose is 10 units of intravenous regular insulin followed immediately by 50 mL of 50% dextrose ($D_{50}W$, 25 g of glucose total); the effect begins in 10–20 min, peaks at 30–60 min, and lasts for 4–6 h. Bolus $D_{50}W$ without insulin is *never* appropriate, given the risk of acutely worsening hyperkalemia due to the osmotic effect of hypertonic glucose. Hypoglycemia is common with insulin plus glucose; hence, this should be followed by an infusion of 10% dextrose at 50–75 mL/h, with close monitoring of plasma glucose concentration. In hyperkalemic patients with glucose concentrations of ≥ 200 –250 mg/dL, insulin should be administered *without* glucose, again with close monitoring of glucose concentrations.

β_2 -Agonists, most commonly albuterol, are effective but underused agents for the acute management of hyperkalemia. Albuterol and insulin with glucose have an additive effect on plasma K⁺ concentration; however, ~20% of patients with ESRD are resistant to the effect of β_2 -agonists; hence, these drugs should not be used without insulin. The recommended dose for inhaled albuterol is 10–20 mg of nebulized albuterol in 4 mL of normal saline, inhaled over 10 min; the effect starts at about 30 min, reaches its peak at about 90 min, and lasts for 2–6 h. Hyperglycemia is a side effect, along with tachycardia. β_2 -Agonists should be used with caution in hyperkalemic patients with known cardiac disease.

Intravenous bicarbonate has no role in the acute treatment of hyperkalemia, but may slowly attenuate hyperkalemia with sustained administration over several hours. It should not be given repeatedly as a hypertonic intravenous bolus of undiluted ampules, given the risk of associated hypernatremia and hypertonicity, but should instead be infused in an isotonic or hypotonic fluid (e.g., 150 milliequivalents of sodium bicarbonate in 1 L of D_5W). In patients with metabolic acidosis, a delayed drop in plasma K⁺ concentration can be seen after 4–6 h of isotonic bicarbonate infusion.

- 3. Removal of potassium.** This is typically accomplished using cation exchange resins, diuretics, and/or dialysis. The cation exchange resin sodium polystyrene sulfonate (SPS) exchanges Na⁺ for K⁺ in the gastrointestinal tract and increases the fecal excretion of

K⁺. The recommended dose of SPS is 15–30 g of powder, almost always given in a premade suspension with 33% sorbitol. The effect of SPS on plasma K⁺ concentration is slow; the full effect may take up to 24 h and usually requires repeated doses every 4–6 h. Intestinal necrosis, typically of the colon or ileum, is a rare but usually fatal complication of SPS. Intestinal necrosis is more common in patients with reduced intestinal motility (e.g., in the postoperative state or after treatment with opioids). The coadministration of SPS with sorbitol appears to increase the risk of intestinal necrosis; however, this complication can also occur with SPS alone, and in animal models, SPS is the causative agent. The low but real risk of intestinal necrosis with SPS, which can sometimes be the only available or appropriate therapy for the removal of potassium, must be weighed against the delayed onset of efficacy. Whenever possible, alternative therapies for the acute management of hyperkalemia (i.e., alternative potassium binders, aggressive redistributive therapy, isotonic bicarbonate infusion, diuretics, and/or hemodialysis) should be used instead of SPS.

Novel intestinal potassium binders have recently become available for the management of hyperkalemia. These agents lack the intestinal toxicity of SPS and are preferred over SPS for the management of hyperkalemia. Patiromer is a nonabsorbed polymer provided as a powder for suspension, which binds K⁺ in exchange for Ca²⁺. In healthy adults, patiromer causes a decrease in urinary potassium, magnesium, and sodium excretion, suggesting the binding of the polymer to these cations in the intestine; notably, a major side effect of the medication is hypomagnesemia. ZS-9 (sodium zirconium cyclosilicate) is an inorganic, nonabsorbable crystalline compound that exchanges both Na⁺ and H⁺ ions in exchange for K⁺ and NH₄⁺ in the intestine. These agents have revolutionized the management of both chronic and acute hyperkalemia. In particular, the availability of safe, well-tolerated potassium binders allows for greater intensity of renin-angiotensin-aldosterone system inhibition in both renal and cardiac disease.

Therapy with intravenous saline may be beneficial in hypovolemic patients with oliguria and decreased distal delivery of Na⁺, with the associated reductions in renal K⁺ excretion. Loop and thiazide diuretics can be used to reduce plasma K⁺ concentration in volume-replete or hypervolemic patients with sufficient renal function for a diuretic response; this may need to be combined with intravenous saline or isotonic bicarbonate to achieve or maintain euolemia.

Hemodialysis is the most effective and reliable method to reduce plasma K⁺ concentration; peritoneal dialysis is considerably less effective. Patients with acute kidney injury require temporary, urgent venous access for hemodialysis, with the attendant risks; in contrast, patients with ESRD or advanced chronic kidney disease may have a preexisting venous access. The amount of K⁺ removed during hemodialysis depends on the relative distribution of K⁺ between ICF and ECF (potentially affected by prior therapy for hyperkalemia), the type and surface area of the dialyzer used, dialysate and blood flow rates, dialysate flow rate, dialysis duration, and the plasma-to-dialysate K⁺ gradient.

FURTHER READING

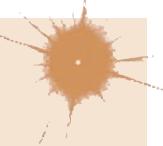
- CHOI M et al: K⁺ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. *Science* 331:768, 2011.
- FENSKE W et al: A copeptin-based approach in the diagnosis of diabetes insipidus. *N Engl J Med* 379:428, 2018.
- GANKAM-KENGNE F et al: Osmotic stress-induced defective glial proteostasis contributes to brain demyelination after hyponatremia treatment. *J Am Soc Nephrol* 28:1802, 2017.
- MOUNT DB: Disorders of potassium balance, in *Brenner and Rector's The Kidney*, 11th ed, ASL Yu et al: (eds). Philadelphia, W.B. Saunders & Company, 2020, pp. 537–579.
- PACKHAM DK et al: Sodium zirconium cyclosilicate in hyperkalemia. *N Engl J Med* 372:222, 2015.

- PERIANAYAGAM A et al: DDAVP is effective in preventing and reversing inadvertent overcorrection of hyponatremia. *Clin J Am Soc Nephrol* 3:331, 2008.
- SCHRIER RW: Decreased effective blood volume in edematous disorders: What does this mean? *J Am Soc Nephrol* 18:2028, 2007.
- SOOD L et al: Hypertonic saline and desmopressin: A simple strategy for safe correction of severe hyponatremia. *Am J Kidney Dis* 61:571, 2013.
- SOUPART A et al: Efficacy and tolerance of urea compared with vaptans for long-term treatment of patients with SIADH. *Clin J Am Soc Nephrol* 7:742, 2012.

54

Hypercalcemia and Hypocalcemia

Sundeep Khosla



The calcium ion plays a critical role in normal cellular function and signaling, regulating diverse physiologic processes such as neuromuscular signaling, cardiac contractility, hormone secretion, and blood coagulation. Thus, extracellular calcium concentrations are maintained within an exquisitely narrow range through a series of feedback mechanisms that involve parathyroid hormone (PTH) and the active vitamin D metabolite 1,25-dihydroxyvitamin D [$1,25(OH)_2D$]. These feedback mechanisms are orchestrated by integrating signals between the parathyroid glands, kidney, intestine, and bone (Fig. 54-1; Chap. 409).

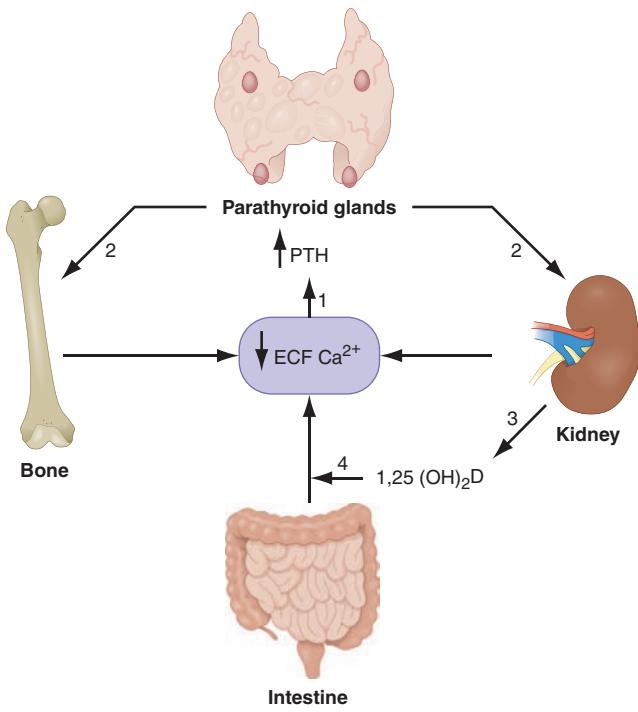


FIGURE 54-1 Feedback mechanisms maintaining extracellular calcium concentrations within a narrow, physiologic range (8.9–10.1 mg/dL [2.2–2.5 mM]). A decrease in extracellular (ECF) calcium (Ca^{2+}) triggers an increase in parathyroid hormone (PTH) secretion (1) via the calcium sensor receptor on parathyroid cells. PTH, in turn, results in increased tubular reabsorption of calcium by the kidney (2) and resorption of calcium from bone (2) and also stimulates renal $1,25(OH)_2D$ production (3). $1,25(OH)_2D$, in turn, acts principally on the intestine to increase calcium absorption (4). Collectively, these homeostatic mechanisms serve to restore serum calcium levels to normal.

Disorders of serum calcium concentration are relatively common and often serve as a harbinger of underlying disease. This chapter provides a brief summary of the approach to patients with altered serum calcium levels. [See Chap. 410 for a detailed discussion of this topic.](#)

HYPERCALCEMIA

■ ETIOLOGY

The causes of hypercalcemia can be understood and classified based on derangements in the normal feedback mechanisms that regulate serum calcium (Table 54-1). Excess PTH production, which is not appropriately suppressed by increased serum calcium concentrations, occurs in primary neoplastic disorders of the parathyroid glands (parathyroid adenomas; hyperplasia; or, rarely, carcinoma) that are associated with increased parathyroid cell mass and impaired feedback inhibition by calcium. Inappropriate PTH secretion for the ambient level of serum calcium also occurs in familial hypocalciuric hypercalcemia (FHH), which is an autosomal dominant syndrome most commonly involving inactivating mutations in the calcium sensor receptor ($CaSR$; FHH type 1), with rare families having mutations in the G_{α_11} protein ($GNA11$; FHH type 2) or the adaptor-related protein complex 2, σ -2 subunit ($AP2S1$; FHH type 3); all of these mutations impair extracellular calcium sensing by the parathyroid glands and the kidneys, leading to inappropriate PTH secretion and increased renal tubular calcium reabsorption. Although PTH secretion by tumors is extremely rare, many solid tumors produce PTH-related peptide (PTHRP), which shares homology with PTH in the first 13 amino acids and binds the PTH receptor, thus mimicking effects of PTH on bone and the kidney. In PTHRP-mediated hypercalcemia of malignancy, PTH levels are suppressed by the high serum calcium levels. Hypercalcemia associated with granulomatous disease (e.g., sarcoidosis) or lymphomas is caused by enhanced conversion of $25(OH)D$ to the potent $1,25(OH)_2D$. In these disorders, $1,25(OH)_2D$ enhances intestinal calcium absorption, resulting in hypercalcemia and suppressed PTH. Disorders that directly increase calcium mobilization from bone, such as hyperthyroidism or osteolytic metastases, also lead to hypercalcemia.

TABLE 54-1 Causes of Hypercalcemia

Excessive PTH production
Primary hyperparathyroidism (adenoma, hyperplasia, rarely carcinoma)
Tertiary hyperparathyroidism (long-term stimulation of PTH secretion in renal insufficiency)
Ectopic PTH secretion (very rare)
FHH
Alterations in $CaSR$ function (lithium therapy)
Hypercalcemia of malignancy
Overproduction of PTHRP (many solid tumors)
Lytic skeletal metastases (breast, myeloma)
Excessive $1,25(OH)_2D$ production
Granulomatous diseases (sarcoidosis, tuberculosis, silicosis)
Lymphomas
Vitamin D intoxication
Primary increase in bone resorption
Hyperthyroidism
Immobilization
Excessive calcium intake
Milk-alkali syndrome
Total parenteral nutrition
Other causes
Endocrine disorders (adrenal insufficiency, pheochromocytoma, VIPoma)
Medications (thiazides, vitamin A, antiestrogens)

Abbreviations: $CaSR$, calcium sensor receptor; FHH, familial hypocalciuric hypercalcemia; PTH, parathyroid hormone; PTHRP, PTH-related peptide.

with suppressed PTH secretion as does exogenous calcium overload, as in milk-alkali syndrome, or total parenteral nutrition with excessive calcium supplementation.

■ CLINICAL MANIFESTATIONS

Mild hypercalcemia (up to 11–11.5 mg/dL) is usually asymptomatic and recognized only on routine calcium measurements. Some patients may complain of vague neuropsychiatric symptoms, including trouble concentrating, personality changes, or depression. Other presenting symptoms may include peptic ulcer disease or nephrolithiasis, and fracture risk may be increased. More severe hypercalcemia (>12–13 mg/dL), particularly if it develops acutely, may result in lethargy, stupor, or coma, as well as gastrointestinal symptoms (nausea, anorexia, constipation, or pancreatitis). Hypercalcemia decreases renal concentrating ability, which may cause polyuria and polydipsia. With long-standing hyperparathyroidism, patients may present with bone pain or pathologic fractures. Finally, hypercalcemia can result in significant electrocardiographic changes, including bradycardia, atrioventricular (AV) block, and short QT interval; changes in serum calcium can be monitored by following the QT interval.

■ DIAGNOSTIC APPROACH

The first step in the diagnostic evaluation of hyper- or hypocalcemia is to ensure that the alteration in serum calcium levels is not due to abnormal albumin concentrations. About 50% of total calcium is ionized, and the rest is bound principally to albumin. Although direct measurements of ionized calcium are possible, they are easily influenced by collection methods and other artifacts; thus, it is generally preferable to measure total calcium and albumin to “correct” the serum calcium. When serum albumin concentrations are reduced, a corrected calcium concentration is calculated by adding 0.2 mM (0.8 mg/dL) to the total calcium level for every decrement in serum albumin of 1.0 g/dL below the reference value of 4.1 g/dL for albumin, and, conversely, for elevations in serum albumin.

A detailed history may provide important clues regarding the etiology of the hypercalcemia (Table 54-1). Chronic hypercalcemia is most commonly caused by primary hyperparathyroidism, as opposed to the second most common etiology of hypercalcemia, an underlying malignancy. The history should include medication use, previous neck surgery, and systemic symptoms suggestive of sarcoidosis or lymphoma.

Once true hypercalcemia is established, the second most important laboratory test in the diagnostic evaluation is a PTH level using a two-site assay for the intact hormone. Increases in PTH are often accompanied by hypophosphatemia. In addition, serum creatinine should be measured to assess renal function; hypercalcemia may impair renal function, and renal clearance of PTH may be altered depending on the fragments detected by the assay. If the PTH level is increased (or “inappropriately normal”) in the setting of elevated calcium and low phosphorus, the diagnosis is almost always primary hyperparathyroidism. Because individuals with FHH may also present with mildly elevated PTH levels and hypercalcemia, this diagnosis should be considered and excluded because parathyroid surgery is ineffective in this condition. A calcium/creatinine clearance ratio (calculated as urine calcium/serum calcium divided by urine creatinine/serum creatinine) of <0.01 is suggestive of FHH, particularly when there is a family history of mild, asymptomatic hypercalcemia. In addition, sequence analysis of the CASR gene is now commonly performed for the definitive diagnosis of FHH, although as noted above, in rare families, FHH may be caused by mutations in the GNA11 or AP2S1 genes, and patients may have to pay out-of-pocket for the genetic analysis. Ectopic PTH secretion is extremely rare.

A suppressed PTH level in the face of hypercalcemia is consistent with non-parathyroid-mediated hypercalcemia, most often due to underlying malignancy. Although a tumor that causes hypercalcemia is generally overt, a PTHrP level may be needed to establish the diagnosis of hypercalcemia of malignancy. Serum 1,25(OH)₂D levels are increased in granulomatous disorders, and clinical evaluation in combination with laboratory testing will generally provide a diagnosis for the various disorders listed in Table 54-1.

TREATMENT

Hypercalcemia

Mild, asymptomatic hypercalcemia does not require immediate therapy, and management should be dictated by the underlying diagnosis. By contrast, significant, symptomatic hypercalcemia usually requires therapeutic intervention independent of the etiology of hypercalcemia. Initial therapy of significant hypercalcemia begins with volume expansion because hypercalcemia invariably leads to dehydration; 4–6 L of intravenous saline may be required over the first 24 h, keeping in mind that underlying comorbidities (e.g., congestive heart failure) may require the use of loop diuretics to enhance sodium and calcium excretion. However, loop diuretics should not be initiated until the volume status has been restored to normal. If there is increased calcium mobilization from bone (as in malignancy or severe hyperparathyroidism), drugs that inhibit bone resorption should be considered. Although salmon calcitonin (4–8 IU/kg intramuscularly or subcutaneously every 6–12 h) is sometimes used, the mainstays of therapy are bisphosphonates, which are potent inhibitors of bone resorption. Zoledronic acid (e.g., 4 mg intravenously over ~30 min) and pamidronate (e.g., 60–90 mg intravenously over 2–4 h) are bisphosphonates that are commonly used for the treatment of hypercalcemia of malignancy in adults. Onset of action is within 1–3 days, with normalization of serum calcium levels occurring in 60–90% of patients. Bisphosphonate infusions may need to be repeated if hypercalcemia relapses. Denosumab (120 mg subcutaneously on days 1, 8, 15, and 29, and then every 4 weeks), an antibody to RANKL, is a potent inhibitor of bone resorption and has been shown to be effective in treating hypercalcemia refractory to bisphosphonates. An alternative to the bisphosphonates or denosumab is gallium nitrate (200 mg/m² intravenously daily for 5 days), which is also effective, but has potential nephrotoxicity. In rare instances, dialysis may be necessary. Finally, although intravenous phosphate chelates calcium and decreases serum calcium levels, this therapy can be toxic because calcium-phosphate complexes may deposit in tissues and cause extensive organ damage.

In patients with 1,25(OH)₂D-mediated hypercalcemia, glucocorticoids are the preferred therapy, as they decrease 1,25(OH)₂D production. Intravenous hydrocortisone (100–300 mg daily) or oral prednisone (40–60 mg daily) for 3–7 days is used most often. Other drugs, such as ketoconazole, chloroquine, and hydroxychloroquine, may also decrease 1,25(OH)₂D production and are used occasionally.

HYPOCALCEMIA

■ ETIOLOGY

The causes of hypocalcemia can be differentiated according to whether serum PTH levels are low (hypoparathyroidism) or high (secondary hyperparathyroidism). Although there are many potential causes of hypocalcemia, impaired PTH production and impaired vitamin D production are the most common etiologies (Table 54-2) (Chap. 410). Because PTH is the main defense against hypocalcemia, disorders associated with deficient PTH production or secretion may be associated with profound, life-threatening hypocalcemia. In adults, hypoparathyroidism most commonly results from inadvertent damage to all four glands during thyroid or parathyroid gland surgery. Hypoparathyroidism is a cardinal feature of autoimmune endocrinopathies (Chap. 388); rarely, it may be associated with infiltrative diseases such as sarcoidosis. Impaired PTH secretion may be secondary to magnesium deficiency or to activating mutations in the CaSR or in the G proteins that mediate CaSR signaling (autosomal dominant hypocalcemia), which suppress PTH, leading to effects that are opposite to those that occur in FHH.

Vitamin D deficiency, impaired 1,25(OH)₂D production (primarily secondary to renal insufficiency), or vitamin D resistance also cause hypocalcemia. However, the degree of hypocalcemia in these disorders is generally not as severe as that seen with hypoparathyroidism because the parathyroids are capable of mounting a compensatory increase in

TABLE 54-2 Causes of Hypocalcemia**Low Parathyroid Hormone Levels (Hypoparathyroidism)**

Parathyroid agenesis
Isolated
DiGeorge's syndrome
Parathyroid destruction
Surgical
Radiation
Infiltration by metastases or systemic diseases
Autoimmune
Reduced parathyroid function
Hypomagnesemia
Autosomal dominant hypocalcemia

High Parathyroid Hormone Levels (Secondary Hyperparathyroidism)

Vitamin D deficiency or impaired 1,25(OH) ₂ D production/action
Nutritional vitamin D deficiency (poor intake or absorption)
Renal insufficiency with impaired 1,25(OH) ₂ D production
Vitamin D resistance, including receptor defects
Parathyroid hormone resistance syndromes
PTH receptor mutations
Pseudohypoparathyroidism (G protein mutations)
Drugs
Calcium chelators
Inhibitors of bone resorption (bisphosphonates, plicamycin)
Altered vitamin D metabolism (phenytoin, ketoconazole)
Miscellaneous causes
Acute pancreatitis
Acute rhabdomyolysis
Hungry bone syndrome after parathyroidectomy
Osteoblastic metastases with marked stimulation of bone formation (prostate cancer)

Abbreviation: PTH, parathyroid hormone.

PTH secretion. Hypocalcemia may also occur in conditions associated with severe tissue injury such as burns, rhabdomyolysis, tumor lysis, or pancreatitis. The cause of hypocalcemia in these settings may include a combination of low albumin, hyperphosphatemia, tissue deposition of calcium, and impaired PTH secretion.

CLINICAL MANIFESTATIONS

Patients with hypocalcemia may be asymptomatic if the decreases in serum calcium are relatively mild and chronic, or they may present with life-threatening complications. Moderate to severe hypocalcemia is associated with paresthesias, usually of the fingers, toes, and circumoral regions, and is caused by increased neuromuscular irritability. On physical examination, a Chvostek's sign (twitching of the circumoral muscles in response to gentle tapping of the facial nerve just anterior to the ear) may be elicited, although it is also present in ~10% of normal individuals. Carpal spasm may be induced by inflation of a blood pressure cuff to 20 mmHg above the patient's systolic blood pressure for 3 min (Trousseau's sign). Severe hypocalcemia can induce seizures, carpopedal spasm, bronchospasm, laryngospasm, and prolongation of the QT interval.

DIAGNOSTIC APPROACH

In addition to measuring serum calcium, it is useful to determine albumin, phosphorus, and magnesium levels. As for the evaluation of

hypercalcemia, determining the PTH level is central to the evaluation of hypocalcemia. A suppressed (or "inappropriately low") PTH level in the setting of hypocalcemia establishes absent or reduced PTH secretion (hypoparathyroidism) as the cause of the hypocalcemia. Further history will often elicit the underlying cause (i.e., parathyroid agenesis vs destruction). By contrast, an elevated PTH level (secondary hyperparathyroidism) should direct attention to the vitamin D axis as the cause of the hypocalcemia. Nutritional vitamin D deficiency is best assessed by obtaining serum 25-hydroxyvitamin D levels, which reflect vitamin D stores. In the setting of renal insufficiency or suspected vitamin D resistance, serum 1,25(OH)₂D levels are informative.

TREATMENT

Hypocalcemia

The approach to treatment depends on the severity of the hypocalcemia, the rapidity with which it develops, and the accompanying complications (e.g., seizures, laryngospasm). Acute, symptomatic hypocalcemia is initially managed with calcium gluconate, 10 mL 10% wt/vol (90 mg or 2.2 mmol) intravenously, diluted in 50 mL of 5% dextrose or 0.9% sodium chloride, given intravenously over 5 min. Continuing hypocalcemia often requires a constant intravenous infusion (typically 10 ampules of calcium gluconate or 900 mg of calcium in 1 L of 5% dextrose or 0.9% sodium chloride administered over 24 h). Accompanying hypomagnesemia, if present, should be treated with appropriate magnesium supplementation.

Chronic hypocalcemia due to hypoparathyroidism is treated with calcium supplements (1000–1500 mg/d elemental calcium in divided doses) and either vitamin D₂ or D₃ (25,000–100,000 U daily) or calcitriol [1,25(OH)₂D, 0.25–2 µg/d]. Other vitamin D metabolites (dihydrotachysterol, alfacalcidiol) are now used less frequently. Importantly, PTH (1-84) (Natpara) is now approved by the Food and Drug Administration for the treatment of refractory hypoparathyroidism, representing an important advance in treatment of these patients. Vitamin D deficiency is best treated using vitamin D supplementation, with the dose depending on the severity of the deficit and the underlying cause. Thus, nutritional vitamin D deficiency generally responds to relatively low doses of vitamin D (50,000 IU, 2–3 times per week for several months), whereas vitamin D deficiency due to malabsorption may require much higher doses (100,000 IU/d or more). The treatment goal is to bring serum calcium into the low normal range and to avoid hypercalcemia, which may lead to nephrolithiasis.

GLOBAL CONSIDERATIONS

In countries with more limited access to health care or screening laboratory testing of serum calcium levels, primary hyperparathyroidism often presents in its severe form with skeletal complications (osteitis fibrosa cystica) in contrast to the asymptomatic form that is common in developed countries. In addition, vitamin D deficiency is paradoxically common in some countries despite extensive sunlight (e.g., India) due to avoidance of sun exposure and poor dietary vitamin D intake.

FURTHER READING

- BILEZIKIAN JP et al: Hyperparathyroidism. *Lancet* 391:168, 2018.
- BRANDI ML et al: Management of hypoparathyroidism: Summary statement and guidelines. *J Clin Endocrinol Metab* 101:2273, 2016.
- HANNAN FM et al: The calcium-sensing receptor in physiology and in calcitropic and noncalcitropic diseases. *Nat Rev Endocrinol* 15:33, 2018.
- MINISOLA S et al: The diagnosis and management of hypercalcemia. *BMJ* 350:h2723, 2015.

NORMAL ACID-BASE HOMEOSTASIS

Systemic arterial pH is maintained between 7.35 and 7.45 by extracellular and intracellular chemical buffering together with respiratory and renal regulatory mechanisms. The control of arterial CO₂ tension (Paco₂) by the central nervous system (CNS) and respiratory system and the control of plasma bicarbonate by the kidneys stabilize the arterial pH by excretion or retention of acid or alkali. The metabolic and respiratory components that regulate systemic pH are described by the Henderson-Hasselbalch equation and solved for pH when the solubility of CO₂ is considered (dissolved CO₂ in mmol/L = 0.03 × Paco₂ in mmHg), at a pK' of 6.1:

$$\text{pH} = \text{pK}' + \log_{10} \frac{[\text{HCO}_3^-]}{\alpha_{\text{CO}_2} \text{PCO}_2}$$

Under most circumstances, CO₂ production and excretion are matched, and the usual steady-state Paco₂ is maintained at ~40 mmHg. Underexcretion of CO₂ produces hypercapnia, and overexcretion causes hypocapnia. Nevertheless, production and excretion are again matched at a new steady-state Paco₂. Therefore, the Paco₂ is regulated primarily by neural respiratory factors and is not subject to regulation by the rate of CO₂ production. Hypercapnia is usually the result of hypoventilation rather than of increased CO₂ production. Increases or decreases in Paco₂ represent derangements of neural respiratory control or are due to compensatory changes in response to a primary alteration in the plasma [HCO₃⁻].

DIAGNOSIS OF GENERAL TYPES OF DISTURBANCES

The most common clinical disturbances are simple acid-base disorders; i.e., metabolic acidosis or alkalosis or respiratory acidosis or alkalosis occurring individually. Recognition of simple acid-base disorders requires appreciation of the limits of physiologic compensation for a primary disturbance.

SIMPLE ACID-BASE DISORDERS

Primary respiratory disturbances (primary changes in Paco₂) invoke compensatory metabolic responses (secondary changes in [HCO₃⁻]), and primary metabolic disturbances elicit predictable compensatory respiratory responses (secondary changes in Paco₂). Physiologic compensation can be predicted from the relationships displayed in **Table 55-1**. In general, with one exception, compensatory responses return the pH toward, but not to, the normal value. Chronic respiratory alkalosis when prolonged is an exception to this rule and may return the pH to a normal value. Metabolic acidosis due to an increase in endogenous acid production (e.g., ketoacidosis or lactic acid acidosis) lowers extracellular fluid [HCO₃⁻] and decreases extracellular pH. This stimulates the medullary chemoreceptors to increase ventilation and to return the ratio of [HCO₃⁻] to Paco₂, and thus pH, toward, but not typically to, the normal value. The degree of respiratory compensation expected in a metabolic acidosis can be predicted from the relationship: Paco₂ = (1.5 × [HCO₃⁻]) + 8 ± 2 (Winter's equation). Thus, applying this equation, a patient with metabolic acidosis and [HCO₃⁻] of 12 mmol/L would be expected to have a Paco₂ of approximately 26 mmHg. In this example, if values for Paco₂ were <24 or >28 mmHg, values that exceed the boundaries for compensation for a simple disorder, a *mixed* disturbance should be recognized (metabolic acidosis plus respiratory alkalosis or metabolic acidosis plus respiratory acidosis, respectively). Compensatory responses for primary metabolic disorders move the Paco₂ in the same direction as the change in [HCO₃⁻], while compensation for primary respiratory disorders moves the [HCO₃⁻] in the same direction as the primary change in Paco₂.

TABLE 55-1 Prediction of Compensatory Responses to Simple Acid-Base Disturbances and Pattern of Changes

DISORDER	PREDICTION OF COMPENSATION	RANGE OF VALUES		
		pH	HCO ₃ ⁻	Paco ₂
Metabolic acidosis	Paco ₂ = (1.5 × HCO ₃ ⁻) + 8 ± 2 or Paco ₂ will ↓ 1.25 mmHg per mmol/L ↓ in [HCO ₃ ⁻] or Paco ₂ = [HCO ₃ ⁻] + 15	Low	Low	Low
Metabolic alkalosis	Paco ₂ will ↑ 0.75 mmHg per mmol/L ↑ in [HCO ₃ ⁻] or Paco ₂ will ↑ 6 mmHg per 10 mmol/L ↑ in [HCO ₃ ⁻] or Paco ₂ = [HCO ₃ ⁻] + 15	High	High	High
Respiratory alkalosis		High	Low	Low
Acute	[HCO ₃ ⁻] will ↓ 0.2 mmol/L per mmHg ↓ in Paco ₂			
Chronic	[HCO ₃ ⁻] will ↓ 0.4 mmol/L per mmHg ↓ in Paco ₂			
Respiratory acidosis		Low	High	High
Acute	[HCO ₃ ⁻] will ↑ 0.1 mmol/L per mmHg ↑ in Paco ₂			
Chronic	[HCO ₃ ⁻] will ↑ 0.4 mmol/L per mmHg ↑ in Paco ₂			

(Table 55-1). Therefore, changes in Paco₂ and [HCO₃⁻] in **opposite directions** (i.e., Paco₂ or [HCO₃⁻] is increased, whereas the other value is decreased) indicate a **mixed acid-base disturbance**. Another way to judge the appropriateness of the response in [HCO₃⁻] or Paco₂ is to use an acid-base nomogram (**Fig. 55-1**). While the shaded areas of the nomogram show the 95% confidence limits for physiologic

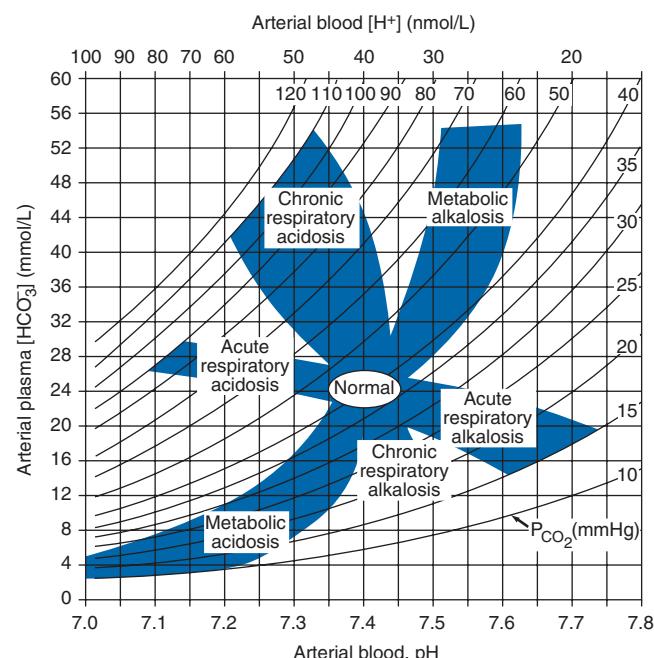


FIGURE 55-1 Acid-base nomogram. Shown are the 95% confidence limits (range of values) of the normal respiratory and metabolic compensations for primary acid-base disturbances. (Reproduced with permission from LL Hamm and TD DuBose Jr, in Alan S.L. Yu, et al (eds): Brenner and Rector's The Kidney, 11th ed. Philadelphia, Elsevier, 2020.)

compensation in simple disturbances, finding acid-base values within the shaded area does not necessarily rule out a mixed disturbance. Imposition of one disorder over another may result in values lying within the area of a third. Thus, the nomogram, while convenient, is not a substitute for the equations in Table 55-1.

MIXED ACID-BASE DISORDERS

Mixed acid-base disorders—defined as independently coexisting disorders, not merely compensatory responses—are often seen in patients in critical care units and can lead to dangerous extremes of pH (**Table 55-2**). The diagnosis of mixed acid-base disorders requires consideration of the anion gap (AG). To be accurate, the AG requires the presence of, or correction to, a normal serum albumin of 4.5 g/dL (see below, “Evaluate the Anion Gap”). If a patient with diabetic ketoacidosis (metabolic acidosis) and a high AG has an independent and concomitant respiratory disorder (e.g., pneumonia), the latter may lead to a superimposed respiratory acidosis or alkalosis and the Paco_2 will deviate from the predicted value for the response to a pure high-AG metabolic acidosis (Table 55-2). Patients with underlying chronic obstructive pulmonary disease may not respond to metabolic acidosis with an appropriate ventilatory response because of insufficient respiratory reserve (Table 55-2). Such imposition of respiratory acidosis on metabolic acidosis can lead to severe acidemia. When metabolic acidosis and metabolic alkalosis coexist in the same patient, the pH may be in the normal range. In this circumstance, it is the presence of an elevated AG (see below) that denotes the presence of a metabolic acidosis. Assuming a normal value for the AG of 10 mmol/L, incongruity in the ΔAG (existing minus normal AG) and the ΔHCO_3^- (normal value of 25 mmol/L minus abnormal HCO_3^- in the patient) indicates the presence of a mixed high-gap acidosis—metabolic alkalosis (see example below). A diabetic patient with ketoacidosis may have acute or chronic kidney failure resulting in a combination of metabolic acidoses from

TABLE 55-2 Examples of Mixed Acid-Base Disorders

Mixed Metabolic and Respiratory

Metabolic acidosis—respiratory alkalosis

Key: High-AG metabolic acidosis; prevailing Paco_2 , *below* predicted value (Table 55-1)

Example: Na^+ , 140; K^+ , 4.0; Cl^- , 106; HCO_3^- , 14; AG, 20; Paco_2 , 24; pH, 7.39 (lactic acidosis, sepsis in ICU)

Metabolic acidosis—respiratory acidosis

Key: High-AG metabolic acidosis; prevailing Paco_2 , *above* predicted value (Table 55-1)

Example: Na^+ , 140; K^+ , 4.0; Cl^- , 102; HCO_3^- , 18; AG, 20; Paco_2 , 38; pH, 7.30 (severe pneumonia, pulmonary edema)

Metabolic alkalosis—respiratory alkalosis

Key: Paco_2 does not increase as predicted; pH higher than expected

Example: Na^+ , 140; K^+ , 4.0; Cl^- , 91; HCO_3^- , 33; AG, 16; Paco_2 , 38; pH, 7.55 (liver disease and diuretics)

Metabolic alkalosis—respiratory acidosis

Key: Paco_2 higher than predicted; pH normal

Example: Na^+ , 140; K^+ , 3.5; Cl^- , 88; HCO_3^- , 42; AG, 10; Paco_2 , 67; pH, 7.42 (COPD on diuretics)

Mixed Metabolic Disorders

Metabolic acidosis—metabolic alkalosis

Key: Only detectable with high-AG acidosis; ΔAG (10) >> ΔHCO_3^-

Example: Na^+ , 140; K^+ , 3.0; Cl^- , 95; HCO_3^- , 25; AG, 20; Paco_2 , 40; pH, 7.42 (uremia with vomiting)

Metabolic acidosis—metabolic acidosis

Key: Mixed high-AG—normal-AG acidosis; ΔHCO_3^- accounted for by combined change in ΔAG and ΔCl^-

Example: Na^+ , 135; K^+ , 3.0; Cl^- , 110; HCO_3^- , 10; AG, 15; Paco_2 , 25; pH, 7.20 (diarrhea and lactic acidosis, toluene toxicity, treatment of diabetic ketoacidosis)

Abbreviations: AG, anion gap; COPD, chronic obstructive pulmonary disease; ICU, intensive care unit.

accumulation of both ketoacids and uremic acids. Patients who have ingested an overdose of drug combinations such as sedatives and salicylates may have mixed disturbances as a result of the acid-base response to the individual drugs (metabolic acidosis mixed with respiratory acidosis or respiratory alkalosis, respectively). Triple acid-base disturbances are more complex. For example, patients with metabolic acidosis due to alcoholic ketoacidosis may develop metabolic alkalosis due to vomiting and superimposed respiratory alkalosis due to the hyperventilation of hepatic dysfunction or alcohol withdrawal.

APPROACH TO THE PATIENT

Acid-Base Disorders

The diagnosis of acid-base disorders follows a stepwise approach (**Table 55-3**). Blood for electrolytes and arterial blood gases should be drawn simultaneously, prior to therapy. An increase in $[\text{HCO}_3^-]$ occurs with either metabolic alkalosis or respiratory acidosis. Conversely, a decrease in $[\text{HCO}_3^-]$ occurs with either metabolic acidosis or respiratory alkalosis. In the determination of arterial blood gases by the clinical laboratory, both pH and Paco_2 are measured, and the $[\text{HCO}_3^-]$ is calculated from the Henderson-Hasselbalch equation. This *calculated* value should be compared with the *measured* $[\text{HCO}_3^-]$ (or total CO_2) on the electrolyte panel. These two values should agree within 2 mmol/L. If they do not, the values may not have been drawn simultaneously, or a laboratory error may be present. After verifying the blood acid-base values, the precise acid-base disorder can then be classified.

EVALUATE THE ANION GAP

Evaluations of acid-base disorders should involve acknowledgement of the AG. The AG is calculated, either by the clinical laboratory or the clinician, as follows: $\text{AG} = \text{Na}^+ - (\text{Cl}^- + \text{HCO}_3^-)$. The value for plasma $[\text{K}^+]$ is typically omitted from the calculation of the AG in the United States. The “normal” value for the AG reported by clinical laboratories has declined with improved methodology for measuring plasma electrolytes and ranges from 6–12 mmol/L, with an average of approximately 10 mmol/L. The unmeasured anions normally present in plasma include anionic proteins (e.g., albumin), phosphate, sulfate, and organic anions. When acid anions, such as acetoacetate and lactate, accumulate in extracellular fluid, the AG increases, causing a **high-AG acidosis**. An increase in the AG is most often due to an increase in unmeasured anions but, less commonly, may be due to a decrease in unmeasured cations (calcium, magnesium, potassium). In addition, the AG may increase with an increase in anionic albumin (e.g., severe dehydration). A decrease in the AG can be due to (1) an increase in unmeasured cations; (2) the addition to the blood of abnormal cations, such as lithium (lithium intoxication) or cationic immunoglobulins (plasma cell dyscrasias); (3) a reduction in the plasma anion albumin concentration (nephrotic syndrome, liver disease, or malabsorption); or (4) hyperviscosity and severe hyperlipidemia, which can lead to an underestimation of sodium and chloride concentrations. Because the normal AG of 10 mmol/L assumes that the serum albumin is normal, if hypoalbuminemia is present, the value for the AG must

TABLE 55-3 Steps in Acid-Base Diagnosis

1. Obtain arterial blood gas (ABG) and electrolytes simultaneously.
2. Compare $[\text{HCO}_3^-]$ on ABG and electrolytes to verify accuracy.
3. Evaluate anion gap (AG); if not normal, correct to albumin concentration of 4.5 g/dL (see text).
4. Know four causes of high-AG acidosis (ketoacidosis, lactic acid acidosis, renal failure, and toxins).
5. Know two causes of hyperchloremic or nongap acidosis (bicarbonate loss from gastrointestinal tract, renal tubular acidosis).
6. Estimate compensatory response (Table 55-1).
7. Compare ΔAG and ΔHCO_3^- .
8. Compare change in $[\text{Cl}^-]$ with change in $[\text{Na}^+]$.

be corrected. For example, for each g/dL of serum albumin below the normal value (4.5 g/dL), 2.5 mmol/L should be added to the reported (uncorrected) AG. Thus, in a patient with a serum albumin of 2.5 g/dL (2 g/dL below the normal value) and an uncorrected AG of 15, the corrected AG is calculated by adding 5 mmol/L ($2.5 \times 2 = 5$; $5 + 15 = \text{corrected AG of } 20 \text{ mmol/L}$). Clinical laboratories do not correct the AG for coexisting hypoalbuminemia and typically report the uncorrected value, requiring the attention of the clinician to the prevailing serum albumin concentration. The clinical disorders that may cause a high-AG acidosis are displayed in Table 55-3.

A high AG is usually due to accumulation of non-chloride-containing acids that contain inorganic (phosphate, sulfate), organic (ketoacids, lactate, uremic organic anions), exogenous (salicylate or ingested toxins with organic acid production), or unidentified anions. The high AG is meaningful even if the $[\text{HCO}_3^-]$ or pH is normal. Simultaneous metabolic acidosis of the high-AG variety plus either chronic respiratory acidosis or metabolic alkalosis represents a situation in which $[\text{HCO}_3^-]$ may be normal or even high (Table 55-3). In cases of high-AG metabolic acidosis, it is valuable to compare the decline in $[\text{HCO}_3^-]$ (ΔHCO_3^- : $25 - \text{patient's } [\text{HCO}_3^-]$) with the increase in the AG (ΔAG : patient's AG – 10).

Similarly, normal values for $[\text{HCO}_3^-]$, Paco_2 , and pH do not ensure the absence of an acid-base disturbance. For instance, an alcoholic who has been vomiting may develop a metabolic alkalosis with a pH of 7.55, Paco_2 of 47 mmHg, $[\text{HCO}_3^-]$ of 40 mmol/L, $[\text{Na}^+]$ of 135, $[\text{Cl}^-]$ of 80, and $[\text{K}^+]$ of 2.8. If such a patient were then to develop a superimposed alcoholic ketoacidosis with a β -hydroxybutyrate concentration of 15 mmol/L, arterial pH would fall to 7.40, the $[\text{HCO}_3^-]$ to 25 mmol/L, and the Paco_2 to 40 mmHg. Although these blood gases are normal, the AG is elevated at 30 mmol/L, indicating a mixed metabolic alkalosis and metabolic acidosis is present. A mixture of high-gap acidosis and metabolic alkalosis is recognized easily by comparing the differences (Δ values) in the normal to prevailing patient values. In this example, the ΔHCO_3^- is 0 (25 – 25 mmol/L), but the ΔAG is 20 (30 – 10 mmol/L). Therefore, 20 mmol/L is unaccounted for in the Δ/Δ value (ΔAG to ΔHCO_3^-).

METABOLIC ACIDOSIS

Metabolic acidosis can occur because of an increase in endogenous acid production (such as lactate and ketoacids), loss of bicarbonate (as in diarrhea), or accumulation of endogenous acids because of inappropriately low excretion of net acid by the kidney (as in chronic kidney disease). Metabolic acidosis has profound effects on the respiratory, cardiac, and nervous systems. The fall in blood pH is accompanied by a characteristic increase in ventilation, especially the tidal volume (Kussmaul respiration). Intrinsic cardiac contractility may be depressed, but inotropic function can be normal because of catecholamine release. Both peripheral arterial vasodilation and central vasoconstriction may be present; the decrease in central and pulmonary vascular compliance predisposes to pulmonary edema with even minimal volume overload. CNS function is depressed, with headache, lethargy, stupor, and, in some cases, even coma. Glucose intolerance may also occur.

There are two major categories of clinical metabolic acidosis: high-AG and non-AG acidosis (Table 55-3 and Table 55-4). The presence of metabolic acidosis, a normal AG, and hyperchloraemia denotes the presence of a non-AG metabolic acidosis.

TABLE 55-4 Causes of High-Anion Gap Metabolic Acidosis

Lactic acidosis	Toxins
Ketoacidosis	Ethylene glycol
Diabetic	Methanol
Alcoholic	Salicylates
Starvation	Propylene glycol
	Pyroglutamic acid (5-oxoproline)
	Renal failure (acute and chronic)

TREATMENT

Metabolic Acidosis

Treatment of metabolic acidosis with alkali should be reserved for severe acidemia except when the patient has no “potential HCO_3^- ” in plasma. The potential $[\text{HCO}_3^-]$ can be estimated from the increment (Δ) in the AG ($\Delta\text{AG} = \text{patient's AG} - 10$), only if the acid anion that has accumulated in plasma is metabolizable (i.e., β -hydroxybutyrate, acetoacetate, and lactate). Conversely, nonmetabolizable anions that may accumulate in advanced-stage chronic kidney disease or after toxin ingestion are not metabolizable and do not represent “potential” HCO_3^- . In patients with acute kidney failure or acute-on-chronic kidney failure, improvement in kidney function after volume resuscitation may improve the serum $[\text{HCO}_3^-]$, but this is a slow and unpredictable process. Consequently, patients with a non-AG acidosis (hyperchloraemic acidosis) or an AG acidosis attributable to a nonmetabolizable anion due to advanced kidney failure (“uremic” acidosis) should receive alkali therapy, either PO (NaHCO_3 tablets or Shohl’s solution) or IV (NaHCO_3), in an amount necessary to slowly increase the plasma $[\text{HCO}_3^-]$ to a target value of 22 mmol/L. Importantly, overcorrection should be avoided.

Bicarbonate therapy in diabetic ketoacidosis (DKA) is reserved for adult patients with severe acidemia ($\text{pH} < 7.00$) and/or evidence of shock. In such circumstances, bicarbonate may be administered IV, as a slow infusion of 50 meq of NaHCO_3 diluted in 300 mL of a saline solution, over 30–45 min, during the initial 1–2 h of therapy. Bolus administration should be avoided. Administration of NaHCO_3 requires careful monitoring of plasma electrolytes during the course of therapy because of the risk for hypokalemia as urine output is established. A reasonable initial goal in DKA is to increase the $[\text{HCO}_3^-]$ to 10–12 mmol/L and the pH to approximately 7.20, but clearly not to increase these values to normal.

HIGH-ANION GAP ACIDOSES

APPROACH TO THE PATIENT

High-Anion Gap Acidoses

There are four principal causes of a high-AG acidosis: (1) lactic acidosis, (2) ketoacidosis, (3) ingested toxins, and (4) acute and chronic kidney failure (Table 55-4). Initial screening to differentiate the high-AG acidoses should include (1) a probe of the history for evidence of drug and toxin ingestion and measurement of arterial blood gas to detect coexistent respiratory alkalosis (salicylates); (2) determination of whether a history of diabetes mellitus is present (DKA); (3) a search for evidence of alcoholism or increased levels of β -hydroxybutyrate (alcoholic ketoacidosis); (4) observation for clinical signs of uremia and determination of the blood urea nitrogen (BUN) and creatinine (uremic acidosis); (5) inspection of the urine for oxalate crystals (ethylene glycol ingestion); and (6) recognition of the numerous clinical settings in which lactate levels may be increased (hypotension, shock, cardiac failure, leukemia, cancer, and drug or toxin ingestion).

Lactic Acidosis An increase in plasma L-lactate may be secondary to poor tissue perfusion (type A)—circulatory insufficiency (shock, cardiac failure), severe anemia, mitochondrial enzyme defects, and inhibitors (carbon monoxide, cyanide)—or to aerobic disorders (type B)—malignancies, nucleoside analogue reverse transcriptase inhibitors in HIV, diabetes mellitus, renal or hepatic failure, thiamine deficiency, severe infections (cholera, malaria), seizures, or drugs/toxins (biguanides, ethanol, and the toxic alcohols: ethylene glycol, methanol, or propylene glycol). Unrecognized bowel ischemia or infarction in a patient with severe atherosclerosis or cardiac decompensation receiving vasopressors is a common cause of lactic acidosis in elderly patients. Pyroglutamic acidemia may occur in critically ill patients.

receiving acetaminophen, which causes depletion of glutathione and accumulation of 5-oxoprolene. D-Lactic acid acidosis, which may be associated with jejunileal bypass, short bowel syndrome, or intestinal obstruction, is due to formation of D-lactate by gut bacteria.

APPROACH TO THE PATIENT

L-Lactic Acid Acidosis

The overarching goal of treatment is to correct the underlying condition that disrupts lactate metabolism; tissue perfusion should be restored when inadequate, but vasoconstrictors should be avoided, or used cautiously, because they may worsen tissue perfusion. Alkali therapy is generally advocated for acute, severe acidemia ($\text{pH} < 7.00$) to improve cardiovascular function. However, NaHCO_3 therapy may paradoxically depress cardiac performance and exacerbate acidosis by enhancing lactate production (HCO_3^- stimulates phosphofructokinase). While the use of alkali in moderate lactic acidosis is controversial, it is generally agreed that attempts to return the pH or $[\text{HCO}_3^-]$ to normal by administration of exogenous NaHCO_3 are deleterious. A reasonable approach with severe acidemia is to infuse sufficient NaHCO_3 to raise arterial pH to no more than 7.2 or the $[\text{HCO}_3^-]$ to no more than 12 mmol/L.

NaHCO_3 therapy can cause fluid overload, hypercapnia, and hypertension because the amount required can be massive when accumulation of lactic acid is relentless. Fluid administration is poorly tolerated, especially in the oliguric patient, when central venoconstriction coexists. If the underlying cause of the lactic acidosis can be remedied, blood lactate will be converted to HCO_3^- and may result in an overshoot alkalosis if exogenous NaHCO_3 has been administered excessively.

Ketoacidosis • DIABETIC KETOACIDOSIS (DKA) This condition is caused by increased fatty acid metabolism and the accumulation of ketoacids (acetooacetate and β -hydroxybutyrate). DKA usually occurs in insulin-dependent diabetes mellitus in association with cessation of insulin or an intercurrent illness such as an infection, gastroenteritis, pancreatitis, or myocardial infarction, which increases insulin requirements temporarily and acutely, and is characterized by hyperglycemia, ketonemia, and a high-AG acidosis. Nevertheless, the plasma glucose may be normal or only slightly elevated in the setting of starvation ketoacidosis or in diabetics receiving antagonists of the proximal tubule sodium-glucose co-transporter 2 (SGLT2). These agents cause glycosuria, an osmotic diuresis, and lower the plasma glucose. Ketoacidosis can occur in patients receiving SGLT2 antagonists for the same reasons as in classical DKA, but the plasma glucose is typically normal or only slightly elevated. The accumulation of ketoacids in plasma accounts for the increment in the AG in both classical DKA and euglycemic DKA. Measurement of urine ketones (by the dipstick nitroprusside reaction) does not detect β -hydroxybutyrate and may underestimate the degree of ketosis (see below). Excretion of ketoacids obligates the excretion of cations, such as Na^+ and K^+ , contributing to volume depletion and Cl^- retention. In some circumstances, a mixed non-AG-high-AG acidosis may occur simultaneously and is recognized when the ΔHCO_3^- exceeds the ΔAG . It should be noted that bicarbonate therapy is rarely necessary in DKA except with extreme acidemia ($\text{pH} < 7.00$) or if the patient is in shock. If administered, NaHCO_3 should be administered in only limited amounts because of the risk for cerebral edema. Patients with DKA are typically volume depleted and require fluid resuscitation with isotonic saline. Volume overexpansion should be avoided, however, because overly aggressive saline administration may cause hyperchloremic acidosis during or following treatment of DKA. Regular insulin should be administered IV as an initial bolus of 0.1 U/kg followed by an infusion of 0.1 U/kg/h until the AG returns to normal; see [Chap. 403](#) for more detail.

ALCOHOLIC KETOACIDOSIS (AKA) AKA is usually associated with chronic alcoholism, binge drinking, vomiting, abdominal pain, poor

nutrition, and volume depletion. The glucose concentration is variable, and acidosis may be severe because of elevated ketones, predominantly β -hydroxybutyrate. The presence of a high-AG acidosis, in the absence of hyperglycemia, in a patient with chronic alcoholism suggests the diagnosis of AKA. Mixed acid-base disorders are common in AKA. Hypoperfusion may enhance lactic acid production (mixed high-AG acidosis), chronic respiratory alkalosis may accompany liver disease (mixed high-AG acidosis and respiratory alkalosis), and metabolic alkalosis can result from vomiting (mixed high-AG acidosis and metabolic alkalosis: ΔAG exceeds ΔHCO_3^-). As the circulation is restored by administration of IV fluids, the preferential accumulation of β -hydroxybutyrate is then shifted to acetooacetate. This explains the common clinical observation of an increasingly positive nitroprusside reaction (ketones) as the circulation is restored. The nitroprusside reaction can detect acetooacetic acid but not β -hydroxybutyrate, so that the degree of ketosis and ketonuria can not only change with therapy, but can be underestimated initially. Therefore, the plasma β -hydroxybutyrate level should be measured. Patients with AKA usually present with relatively normal renal function, as opposed to DKA, where renal function is often compromised because of volume depletion (osmotic diuresis) or diabetic nephropathy. The AKA patient with normal renal function may excrete relatively large quantities of ketoacids and retain Cl^- and, therefore, may have a mixed high-AG-non-AG metabolic acidosis (ΔHCO_3^- exceeds ΔAG).

TREATMENT

Alcoholic Ketoacidosis

Extracellular fluid deficits almost always accompany AKA and should be repaired by IV administration, initially, of saline and glucose (5% dextrose in 0.9% NaCl). Hypophosphatemia, hypokalemia, and hypomagnesemia may coexist and should be monitored carefully and corrected when indicated. Hypophosphatemia may emerge 12–24 h after admission, exacerbated by glucose infusion, and, if severe, may induce marked muscle weakness, hemolysis, rhabdomyolysis, or respiratory arrest. Upper gastrointestinal hemorrhage, pancreatitis, and pneumonia may accompany this disorder.

Drug- and Toxin-Induced Acidosis • SALICYLATES (See also Chap. 458) Salicylate intoxication in adults usually causes respiratory alkalosis or a mixture of high-AG metabolic acidosis and respiratory alkalosis. Only a portion of the AG is due to salicylates. Lactic acid production is also often increased.

TREATMENT

Salicylate-Induced Acidosis

Vigorous gastric lavage with isotonic saline (not NaHCO_3) should be initiated immediately. All patients should receive at least one round of activated charcoal per nasogastric tube (1 g/kg up to 50 g). To facilitate excretion of salicylate in the acidotic patient, IV NaHCO_3 is administered in amounts adequate to alkalinize the urine (urine $\text{pH} > 7.5$) and to maintain urine output. Raising urine pH from 6.5 to 7.5 increases salicylate clearance fivefold. Patients with coexisting respiratory alkalosis should also receive NaHCO_3 cautiously to avoid excessive alkalemia. Acetazolamide may be administered in the face of alkalemia, when an alkaline diuresis cannot be achieved, or to ameliorate volume overload associated with NaHCO_3 administration. Acetazolamide may cause systemic metabolic acidosis if the excreted HCO_3^- is not replaced, a circumstance that can markedly reduce salicylate clearance. **Hypokalemia should be anticipated** with vigorous bicarbonate therapy and should be treated promptly and aggressively. Glucose-containing fluids should be administered because of the danger of hypoglycemia. Excessive insensible fluid losses may cause severe volume depletion and

hypernatremia. If renal failure prevents rapid clearance of salicylate, hemodialysis should be performed against a standard bicarbonate dialysate ($[HCO_3^-] = 30\text{--}35 \text{ meq/L}$).

ALCOHOLS Under most physiologic conditions, sodium, urea, and glucose generate the osmotic pressure of blood. Plasma osmolality is calculated according to the following expression: $P_{\text{osm}} = 2\text{Na}^+ + \text{Glu} + \text{BUN}$ (all in mmol/L), or, using conventional laboratory values in which glucose and BUN are expressed in mg/dL: $P_{\text{osm}} = 2\text{Na}^+ + \text{Glu}/18 + \text{BUN}/2.8$. The calculated and determined osmolality should agree within 10–15 mmol/kg H₂O. When the measured osmolality exceeds the calculated osmolality by >10–15 mmol/kg H₂O, one of two circumstances prevails. Either the serum sodium is spuriously low, as with hyperlipidemia or hyperproteinemia (pseudohyponatremia), or osmolytes other than sodium salts, glucose, or urea have accumulated in plasma. Examples of such osmolytes include mannitol, radiocontrast media, ethanol, isopropyl alcohol, ethylene glycol, propylene glycol, methanol, and acetone. In this situation, the difference between the calculated osmolality and the measured osmolality (*osmolar gap*) is proportional to the concentration of the unmeasured solute. With an appropriate clinical history and index of suspicion, identification of an osmolar gap is helpful in identifying the presence of toxic alcohol-associated AG acidosis. Three alcohols may cause fatal intoxications: ethylene glycol, methanol, and isopropyl alcohol. All cause an elevated osmolal gap, but only the first two cause a high-AG acidosis. Isopropyl alcohol ingestion does not typically elevate the AG unless extreme overdose causes hypotension and lactic acid acidosis.

ETHYLENE GLYCOL (See also Chap. 458) Ethylene glycol (EG) (commonly used in antifreeze, but also in brake fluid and windshield washer fluid deicers) is metabolized by alcohol dehydrogenase, and ingestion of EG leads to a metabolic acidosis and severe damage to the CNS, heart, lungs, and kidneys. The combination of both a high AG and osmolar gap is highly suspicious for EG or methanol intoxication. The combination of a high AG and high osmolar gap in a patient suspected of EG ingestion should be taken as evidence of EG toxicity prior to measurement of EG levels, and treatment should not be delayed. The osmolar gap may be elevated earlier than the AG, and as the osmolar gap declines, the AG increases. The increased AG and osmolar gap in EG intoxication are attributable to EG and its metabolites, glycolate, oxalate, and other organic acids. Lactic acid production increases secondary to inhibition of the tricarboxylic acid cycle and altered intracellular redox state and may contribute to the high AG. Acute tubule injury is caused initially by glycolate and later is amplified by tubule obstruction from oxalate crystals.

TREATMENT

Ethylene Glycol Intoxication

This includes the prompt institution of IV isotonic fluids, thiamine and pyridoxine supplements, fomepizole, and usually, hemodialysis. Both fomepizole and ethanol compete with EG for metabolism by alcohol dehydrogenase. Fomepizole (4-methylpyrazole; 15 mg/kg IV over 30 min as a loading dose, then 10 mg/kg for four doses every 12 h) is the agent of choice and offers the advantages of a predictable decline in EG levels without excessive obtundation, as seen during ethyl alcohol infusion. Fomepizole should be continued until blood pH is normal or the osmolar gap is <10 mOsm/kg H₂O. Hemodialysis is indicated when the arterial pH is <7.3, a high-AG acidosis is present, the osmolar gap exceeds 20 mOsm/kg H₂O, or there is evidence of end organ damage such as CNS manifestations and kidney failure.

METHANOL (See also Chap. 458) The ingestion of methanol (wood alcohol) causes metabolic acidosis, and its metabolites formaldehyde and formic acid cause severe optic nerve and CNS damage. Lactic acid, ketoacids, and other unidentified organic acids may contribute to the acidosis. Due to its low molecular mass (32 Da), an osmolar gap is present and may precede the elevation of the AG.

TREATMENT

Methanol Intoxication

Treatment of methanol intoxication is similar to that for EG intoxication, including general supportive measures, fomepizole, and hemodialysis.

PROPYLENE GLYCOL Propylene glycol is the vehicle used in IV administration of diazepam, lorazepam, phenobarbital, nitroglycerine, etomidate, enoximone, and phenytoin. Propylene glycol is generally safe for limited use in these IV preparations, but toxicity has been reported in the setting of the intensive care unit in patients receiving frequent or continuous therapy, where the propylene glycol vehicle may accumulate in the plasma. This form of high-gap acidosis should be considered in patients with unexplained high-gap acidosis, hyperosmolality, and clinical deterioration, especially in the setting of treatment for alcohol withdrawal. Propylene glycol, like EG and methanol, is metabolized by alcohol dehydrogenase. With intoxication by propylene glycol, the first response is to stop the offending infusion. Additionally, fomepizole should also be administered in acidotic patients.

ISOPROPYL ALCOHOL Ingested isopropanol is absorbed rapidly and may be fatal when as little as 150 mL of rubbing alcohol, solvent, or deicer is consumed. A plasma level >400 mg/dL is life-threatening. Isopropyl alcohol is metabolized by alcohol dehydrogenase to acetone. The characteristic features differ significantly from EG and methanol intoxication in that the parent compound, not the metabolites, causes toxicity, and a high-AG acidosis is *not* present because acetone is rapidly excreted. Both isopropyl alcohol and acetone increase the osmolar gap, and hypoglycemia is common. Alternative diagnoses should be considered if the patient does not improve significantly within a few hours. Patients with hemodynamic instability with plasma levels above 400 mg/dL should be considered for hemodialysis.

TREATMENT

Isopropyl Alcohol Toxicity

Isopropanol alcohol toxicity is treated by supportive therapy, IV fluids, pressors, ventilatory support if needed, and acute hemodialysis for prolonged coma, hemodynamic instability, or levels >400 mg/dL.

PYROGLUTAMIC ACID Acetaminophen-induced high-AG metabolic acidosis is uncommon but is recognized in either patients with acetaminophen overdose or malnourished or critically ill patients receiving acetaminophen in typical dosage. 5-Oxoproline accumulation after acetaminophen should be suspected in the setting of an unexplained high-AG acidosis without elevation of the osmolar gap in patients receiving acetaminophen. The first step in treatment is to immediately discontinue acetaminophen. Additionally, sodium bicarbonate IV should be given. Although N-acetylcysteine has been suggested, it is not proven that it hastens the metabolism of 5-oxoproline by increasing intracellular glutathione concentrations in this setting, as assumed.

Chronic Kidney Disease (See also Chap. 311) The hyperchloremic acidosis of moderate chronic kidney disease (CKD; stage 3) is eventually converted to the high-AG acidosis of advanced renal failure (stages 4 and 5 CKD). Poor filtration and reabsorption of organic anions contribute to the pathogenesis. As renal disease progresses, the number of functioning nephrons eventually becomes insufficient to keep pace with net acid production. Uremic acidosis in advanced CKD is characterized, therefore, by a reduced rate of NH₄⁺ production and excretion. Alkaline salts from bone buffer the acid retained in CKD. Despite significant retention of acid (up to 20 mmol/d), the serum [HCO₃⁻] does not typically decrease further, indicating participation of buffers outside the extracellular compartment. Therefore, the trade-off in untreated chronic metabolic acidosis of CKD stages 3 and 4 is significant loss of bone mass due to reduction in bone calcium carbonate.

Chronic acidosis also contributes significantly to muscle wasting and disability in advancing CKD.

TREATMENT

Metabolic Acidosis of Chronic Kidney Disease

Because of the association of metabolic acidosis in advanced CKD with muscle catabolism, bone disease, and more rapid progression of CKD, both the “uremic acidosis” of end-stage renal disease and the non-AG metabolic acidosis of stages 3 and 4 CKD require oral alkali replacement to increase and maintain the $[HCO_3^-]$ to a value >22 mmol/L. This can be accomplished with relatively modest amounts of alkali (1.0–1.5 mmol/kg body weight per day) and has been shown to slow the progression of CKD. Either NaHCO₃ tablets (650-mg tablets contain 7.8 meq) or oral sodium citrate (Shohl's solution) is effective. Moreover, addition of fruits and vegetables (citrate) to the diet may increase the plasma $[HCO_3^-]$ and slow progression.

NON-ANION GAP METABOLIC ACIDOSES

Alkali can be lost from the gastrointestinal tract as a result of diarrhea or from the kidneys due to renal tubular abnormalities (e.g., renal tubular acidosis [RTA]). In these disorders (Table 55-5), reciprocal

TABLE 55-5 Causes of Non-Anion Gap Acidosis

- I. Gastrointestinal bicarbonate loss
 - A. Diarrhea
 - B. External pancreatic or small-bowel drainage
 - C. Ureterosigmoidostomy, jejunal loop, ileal loop
 - D. Drugs
 - 1. Calcium chloride (acidifying agent)
 - 2. Magnesium sulfate (diarrhea)
 - 3. Cholestyramine (bile acid diarrhea)
- II. Renal acidosis
 - A. Hypokalemia
 - 1. Proximal RTA (type 2)
 - Drug-induced: acetazolamide, topiramate
 - 2. Distal (classic) RTA (type 1)
 - Drug-induced: amphotericin B, ifosfamide
 - B. Hyperkalemia
 - 1. Generalized distal nephron dysfunction (type 4 RTA)
 - a. Selective aldosterone deficiency
 - b. Mineralocorticoid resistance (PHA I, autosomal dominant)
 - c. Voltage defect (PHA I, autosomal recessive, and PHA II)
 - d. Hyporeninemic hypoaldosteronism
 - e. Tubulointerstitial disease
 - C. Normokalemia
 - 1. Chronic progressive kidney disease
- III. Drug-induced hyperkalemia (with renal insufficiency)
 - A. Potassium-sparing diuretics (amiloride, triamterene, spironolactone, eplerenone)
 - B. Trimethoprim
 - C. Pentamidine
 - D. ACE-Is and ARBs
 - E. Nonsteroidal anti-inflammatory drugs
 - F. Calcineurin inhibitors
 - G. Heparin in critically ill patients
- IV. Other
 - A. Acid loads (ammonium chloride, hyperalimentation)
 - B. Loss of potential bicarbonate: ketosis with ketone excretion
 - C. Expansion acidosis (rapid saline administration)
 - D. Hippurate
 - E. Cation exchange resins

Abbreviations: ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; PHA, pseudohypoaldosteronism; RTA, renal tubular acidosis.

changes in $[Cl^-]$ and $[HCO_3^-]$ result in a normal AG. In non-AG acidosis, therefore, the increase in $[Cl^-]$ above the normal value approximates the decrease in $[HCO_3^-]$. The absence of such a relationship suggests a mixed disturbance.

Stool contains a higher concentration of HCO_3^- and decomposed HCO_3^- than plasma so that metabolic acidosis develops in diarrhea. Instead of an acid urine pH (as anticipated with systemic acidosis), urine pH is usually >6 because metabolic acidosis and hypokalemia increase renal synthesis and excretion of NH_4^+ , thus providing a urinary buffer that increases urine pH. Metabolic acidosis due to gastrointestinal losses with a high urine pH can be differentiated from RTA because urinary NH_4^+ excretion is typically low in RTA and high with diarrhea. Urinary NH_4^+ levels are not routinely measured by clinical laboratories but can be estimated by calculating the urine anion gap (UAG): $UAG = [Na^+ + K^+]_u - [Cl^-]_u$. When $[Cl^-]_u > [Na^+ + K^+]_u$, the UAG is negative by definition. This suggests that the urine ammonium level is appropriately increased, suggesting an extrarenal cause of the acidosis. Conversely, when the UAG is positive, the urine ammonium level is predictably low, suggesting a renal tubular origin of the acidosis. Recent studies have shown a poor correlation between the UAG and the measured urine ammonium, thus calling the estimation of urine ammonium by calculation of the UAG into question. Therefore, clinical laboratories should be encouraged to measure urine ammonium by adaptation of automated plasma ammonium assays, using the enzymatic method, if the urine sample is diluted 1:200 in normal saline.

Proximal RTA (type 2 RTA) (**Chap. 315**) is most often due to generalized proximal tubular dysfunction manifested by glycosuria, generalized aminoaciduria, and phosphaturia (Fanconi syndrome). When the plasma $[HCO_3^-]$ is low, the urine pH is acid ($pH < 5.5$) but exceeds 5.5 with alkali therapy. The fractional excretion of $[HCO_3^-]$ may exceed 10–15% when the serum HCO_3^- is >20 mmol/L. Because of the defect in HCO_3^- reabsorption by the proximal tubule, therapy with NaHCO₃ will enhance delivery of HCO_3^- to the distal nephron and enhance renal potassium secretion, thereby causing hypokalemia.

The typical findings in acquired or inherited forms of **classic distal RTA** (type 1 RTA) include hypokalemia, a non-AG metabolic acidosis, low urinary NH_4^+ excretion (positive UAG, low urine $[NH_4^+]$), and inappropriately high urine pH ($pH > 5.5$). Most patients have hypocitraturia and hypercalciuria; nephrolithiasis, nephrocalcinosis, and bone disease are common. In **generalized distal RTA** (type 4 RTA), hyperkalemia is disproportionate to the reduction in glomerular filtration rate (GFR) because of coexisting dysfunction of potassium and acid secretion. Urinary ammonium excretion is invariably depressed, and kidney function may be compromised secondary to diabetic nephropathy, obstructive uropathy, or chronic tubulointerstitial disease.

Hyporeninemic hypoaldosteronism typically presents as a non-AG metabolic acidosis in older adults with diabetes mellitus or tubulointerstitial disease and CKD (estimated GFR 20–50 mL/min) with hyperkalemia ($[K^+]$ 5.2–6.0 mmol/L), concurrent hypertension, and congestive heart failure. Both the metabolic acidosis and the hyperkalemia are out of proportion to impairment in GFR. Nonsteroidal anti-inflammatory drugs, trimethoprim, pentamidine, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin receptor blockers (ARBs) can also increase the risk for hyperkalemia and a non-AG metabolic acidosis in patients with CKD, especially from diabetic nephropathy (Table 55-5).

TREATMENT

Non-Anion Gap Metabolic Acidoses

For non-AG acidosis due to gastrointestinal losses of bicarbonate, NaHCO₃ may be administered intravenously or orally, as determined by the severity of both the acidosis and the accompanying volume depletion. Proximal RTA is the most challenging of the RTAs to treat if the goal is to restore the serum $[HCO_3^-]$ to normal because administration of oral alkali increases urinary excretion of bicarbonate and potassium. In patients with proximal RTA (type 2), potassium

administration is typically required. An oral solution of sodium and potassium citrate (citric acid 334 mg, sodium citrate 500 mg, and potassium citrate 550 mg per 5 mL) may be prescribed for this purpose (Virtrate or Cytra-3). In classical distal RTA (type 1), hypokalemia should be corrected first. When accomplished, alkali therapy with either sodium citrate (Shohl's solution) or NaHCO₃ tablets (650-mg tablets contain 7.8 meq) should be initiated to correct and maintain the serum [HCO₃⁻] in the range of 24–26 meq/L. Type 1 RTA patients typically respond to chronic alkali therapy readily, and the benefits of adequate alkali therapy include a decrease in the frequency of nephrolithiasis, improvement in bone density, resumption of normal growth patterns in children, and preservation of kidney function in both adults and children. For type 4 RTA, attention must be paid to the dual goals of correction of the metabolic acidosis, using the same approach as for classical distal renal tubular acidosis (type 1 RTA), and also correction of the plasma [K⁺]. Restoration of normokalemia increases urinary net acid excretion and consequently can greatly improve the metabolic acidosis. Chronic administration of oral sodium polystyrene sulfonate (15 g of powder prepared as an oral solution, without sorbitol, once daily 2–3 times per week) is sometimes used but is unpalatable, and patient compliance is low. The nonabsorbed, calcium-potassium cation exchange polymer, patiromer, may be considered for type 4 RTA patients with hyperkalemia because it is more palatable. It is administered as 8.4-g packets of powder for suspension PO twice daily with dose adjustment at weekly intervals, based on the plasma [K⁺], not to exceed 25.2 g/d. Additionally, the diet should be low in potassium-containing foods or supplements (salt substitute), all potassium-retaining medications should be discontinued, and a loop diuretic may be administered. Finally, patients with documented isolated hypoaldosteronism should receive fludrocortisone, but the dose varies with the cause of the hormone deficiency. This agent should be administered very cautiously and in combination with furosemide in patients with edema and hypertension because of possible aggravation of these conditions.

METABOLIC ALKALOSIS

Metabolic alkalosis is established by an elevated arterial pH, an increase in the serum [HCO₃⁻], and an increase in Paco₂ as a result of compensatory alveolar hypoventilation (Table 55-1). It is often accompanied by hypochloremia and hypokalemia. The elevation in arterial pH establishes the diagnosis because pH is decreased in respiratory acidosis, even though both have an elevated Paco₂. Metabolic alkalosis frequently occurs as a mixed acid-base disorder in association with either respiratory acidosis, respiratory alkalosis, or metabolic acidosis.

ETIOLOGY AND PATHOGENESIS

Metabolic alkalosis occurs as a result of net gain of [HCO₃⁻] or loss of nonvolatile acid (usually HCl by vomiting) from the extracellular fluid. When vomiting causes loss of HCl from the stomach, HCO₃⁻ secretion cannot be initiated in the small bowel, and thus, HCO₃⁻ is retained in the extracellular fluid. Thus, vomiting or nasogastric suction is an example of the *generation stage* of metabolic alkalosis, in which the loss of acid typically causes alkalosis. Upon cessation of vomiting, the *maintenance stage* ensues because secondary factors prevent the kidneys from excreting HCO₃⁻ appropriately.

Maintenance of metabolic alkalosis, therefore, represents a failure of the kidneys to eliminate excess HCO₃⁻ from the extracellular compartment. The kidneys will retain, rather than excrete, the excess alkali and maintain the alkalosis if (1) volume deficiency, chloride deficiency, and K⁺ deficiency exist in combination with a reduced GFR (associated with a low urine [Cl⁻]) or (2) hypokalemia exists because of autonomous hyperaldosteronism (normal urine [Cl⁻]). In the first example, saline-responsive metabolic alkalosis is corrected by extracellular fluid volume (ECFV) restoration (IV administration of NaCl and KCl), whereas, in the latter, it may be necessary to repair the alkalosis by pharmacologic or surgical intervention, not with saline administration (saline-unresponsive metabolic alkalosis).

TABLE 55-6 Causes of Metabolic Alkalosis

- I. Exogenous HCO₃⁻ loads
 - A. Acute alkali administration
 - B. Milk-alkali syndrome
- II. Effective ECFV contraction, normotension, K⁺ deficiency, and secondary hyperreninemic hyperaldosteronism
 - A. Gastrointestinal origin
 - 1. Vomiting
 - 2. Gastric aspiration
 - 3. Congenital chloridorrhea
 - 4. Gastrocystoplasty
 - 5. Villous adenoma
 - B. Renal origin
 - 1. Diuretic use (thiazides and loop diuretics)
 - 2. Posthypcapnic state
 - 3. Hypercalcemia/hypoparathyroidism
 - 4. Recovery from lactic acidosis or ketoacidosis
 - 5. Nonreabsorbable anions including penicillin, carbenicillin
 - 6. Mg²⁺ deficiency
 - 7. K⁺ depletion
 - 8. Bartter's syndrome (loss-of-function mutations of transporters and ion channels in TALH)
 - 9. Gitelman's syndrome (loss-of-function mutation of Na⁺-Cl⁻ cotransporter in DCT)
- III. ECFV expansion, hypertension, K⁺ deficiency, and mineralocorticoid excess
 - A. High renin
 - 1. Renal artery stenosis
 - 2. Accelerated hypertension
 - 3. Renin-secreting tumor
 - 4. Estrogen therapy
 - B. Low renin
 - 1. Primary aldosteronism
 - a. Adenoma
 - b. Hyperplasia
 - c. Carcinoma
 - 2. Adrenal enzyme defects
 - a. 11β-Hydroxylase deficiency
 - b. 17α-Hydroxylase deficiency
 - 3. Cushing's syndrome or disease
 - 4. Other
 - a. Licorice
 - b. Carbenoxolone
 - c. Chewer's tobacco
- IV. Gain-of-function mutation of sodium channel in DCT with ECFV expansion, hypertension, K⁺ deficiency, and hyporeninemic-hypoaldosteronism
 - A. Liddle's syndrome

Abbreviations: DCT, distal convoluted tubule; ECFV, extracellular fluid volume; TALH, thick ascending limb of Henle's loop.

DIFFERENTIAL DIAGNOSIS

To establish the cause of metabolic alkalosis (Table 55-6), it is necessary to assess the status of the ECFV, the recumbent and upright blood pressure (to determine if orthostasis is present), the serum [K⁺], the urine [Cl⁻], and in some circumstances, the renin-aldosterone system. For example, the presence of chronic hypertension and chronic hypokalemia in an alkalotic patient suggests either mineralocorticoid excess or that the hypertensive patient is receiving diuretics. Low plasma renin activity and values for urine [Cl⁻] >20 meq/L in a patient who is not taking diuretics suggest primary mineralocorticoid excess. The combination of hypokalemia and alkalosis in a normotensive, nonedematous patient can be due to Bartter's or Gitelman's syndrome, magnesium deficiency, vomiting, exogenous alkali, or diuretic ingestion. Measurement of urine electrolytes (especially the urine [Cl⁻]) and screening of the urine for diuretics are recommended. If the urine is alkaline, with an elevated [Na⁺]_u and [K⁺]_u but low [Cl⁻]_u, the diagnosis is usually

either vomiting (overt or surreptitious) or alkali ingestion. If the urine is relatively acid with low concentrations of Na^+ , K^+ , and Cl^- , the most likely possibilities are prior vomiting, the posthypercapnic state, or prior diuretic ingestion. If the urine sodium, potassium, and chloride concentrations are not depressed, magnesium deficiency, Bartter's or Gitelman's syndrome, or current diuretic ingestion should be considered. Bartter's syndrome is distinguished from Gitelman's syndrome by the presence of hypocalciuria in the latter disorder.

Alkali Administration Chronic administration of alkali to individuals with normal renal function rarely causes alkalosis. However, in patients with coexistent hemodynamic disturbances associated with effective ECFV depletion (e.g., heart failure), alkalosis can develop because of diminished capacity to excrete HCO_3^- or enhanced reabsorption of HCO_3^- . Such patients include those who receive NaHCO_3 (PO or IV), citrate loads IV (transfusions of whole blood, or therapeutic apheresis), or antacids plus cation-exchange resins (aluminum hydroxide and sodium polystyrene sulfonate). Nursing home patients receiving enteral tube feedings have a higher incidence of metabolic alkalosis than nursing home patients receiving regular diets.

METABOLIC ALKALOSIS ASSOCIATED WITH ECFV CONTRACTION, K^+ DEPLETION, AND SECONDARY HYPERRENINEMIC HYPERALDOSTERONISM

Gastrointestinal Origin Gastrointestinal loss of H^+ from vomiting or gastric aspiration causes simultaneous addition of HCO_3^- into the extracellular fluid. During active vomiting, the filtered load of bicarbonate reaching the kidneys is acutely increased and will exceed the reabsorptive capacity of the proximal tubule for HCO_3^- absorption. Subsequently, enhanced delivery of HCO_3^- to the distal nephron, where the capacity for HCO_3^- reabsorption is lower, will result in excretion of alkaline urine that stimulates potassium secretion. When vomiting ceases, the persistence of volume, potassium, and chloride depletion triggers maintenance of the alkalosis because these conditions promote HCO_3^- reabsorption. Correction of the contracted ECFV with NaCl and repair of K^+ deficits with KCl corrects the acid-base disorder by restoring the ability of the kidney to excrete the excess bicarbonate.

Renal Origin • DIURETICS (See also Chap. 258) Diuretics such as thiazides and loop diuretics (furosemide, bumetanide, torsemide) increase excretion of salt and acutely diminish the ECFV without altering the total body bicarbonate content. The serum $[\text{HCO}_3^-]$ increases because the reduced ECFV "contracts" around the $[\text{HCO}_3^-]$ in plasma (contraction alkalosis). The chronic administration of diuretics tends to generate an alkalosis by increasing distal salt delivery so that both K^+ and H^+ secretion are stimulated. The alkalosis is maintained by persistence of the contraction of the ECFV, secondary hyperaldosteronism, K^+ deficiency, and the direct effect of the diuretic (as long as diuretic administration continues). Discontinuing the diuretic and providing isotonic saline to correct the ECFV deficit will repair the alkalosis.

SOLUTE LOSING DISORDERS: BARTTER'S SYNDROME AND GITELMAN'S SYNDROME See Chap. 315.

NON-REABSORBABLE ANIONS AND MAGNESIUM DEFICIENCY Administration of large quantities of the penicillin derivatives carbencillin or ticarcillin cause their non-reabsorbable anions to appear in the distal tubule. This increases the transepithelial potential difference in the collecting tubule and thereby enhances H^+ and K^+ secretion. Mg^{2+} deficiency may occur with chronic administration of thiazide diuretics, alcoholism, and malnutrition, and in Gitelman's syndrome, it potentiates the development of hypokalemic alkalosis by enhancing distal acidification through stimulation of renin and hence aldosterone secretion.

POTASSIUM DEPLETION Chronic K^+ depletion as a result of extreme dietary potassium insufficiency, diuretics, or alcohol abuse may initiate metabolic alkalosis by increasing urinary net acid excretion. The

renal generation of NH_4^+ (ammoniagenesis) is upregulated directly by hypokalemia. Chronic K^+ deficiency also upregulates the H^+ , K^+ -ATPases in the distal tubule and collecting duct to increase K^+ absorption while simultaneously increasing H^+ secretion. Alkalosis associated with severe K^+ depletion is resistant to salt administration, but repair of the K^+ deficiency corrects the alkalosis. Potassium depletion often occurs concurrent with magnesium deficiency in alcoholics with malnutrition.

AFTER TREATMENT OF LACTIC ACIDOSIS OR KETOACIDOSIS When an underlying stimulus for the generation of lactic acid or ketoacid is corrected, such as shock or severe volume depletion by volume restoration, or with insulin therapy, the lactate or ketones are metabolized to yield an equivalent amount of HCO_3^- . Exogenous sources of HCO_3^- will be additive to that amount generated by organic anion metabolism and may create a surfeit of HCO_3^- ("rebound alkalosis").

POSTHYPERCAPNIA Prolonged CO_2 retention with chronic respiratory acidosis enhances renal HCO_3^- absorption and the generation of new HCO_3^- (increased net acid excretion). Metabolic alkalosis results from the persistently elevated $[\text{HCO}_3^-]$ when the elevated Paco_2 is abruptly returned toward normal.

METABOLIC ALKALOSIS ASSOCIATED WITH ECFV EXPANSION, HYPERTENSION, AND MINERALOCORTICOID EXCESS

Increased aldosterone levels may be the result of autonomous primary adrenal overproduction or of secondary aldosterone release due to renal overproduction of renin. Mineralocorticoid excess increases net acid excretion and may result in metabolic alkalosis, which is typically exacerbated by associated K^+ deficiency. Salt retention and hypertension are due to upregulation of the epithelial Na^+ channel (ENaC) in the collecting tubule in response to aldosterone. The kaliuresis persists because of mineralocorticoid excess and stimulation of ENaC, causing an increase in transepithelial voltage, which enhances K^+ excretion. Persistent K^+ depletion may cause polydipsia and polyuria.

Liddle's syndrome (Chap. 315) results from an inherited gain-of-function mutation of genes that regulate the collecting duct Na^+ channel, ENaC. This rare monogenic form of hypertension is the result of volume expansion that secondarily suppresses aldosterone elaboration. Patients typically present with hypertension, hypokalemia, and metabolic alkalosis.

Symptoms With metabolic alkalosis, changes in CNS and peripheral nervous system function are similar to those of hypocalcemia (Chap. 409); symptoms include mental confusion; obtundation; and a predisposition to seizures, paresthesias, muscular cramping, tetany, aggravation of arrhythmias, and hypoxemia in chronic obstructive pulmonary disease. Related electrolyte abnormalities include hypokalemia and hypophosphatemia.

TREATMENT

Metabolic Alkalosis

The first goal of therapy is to correct the underlying stimulus for HCO_3^- generation. If primary aldosteronism or Cushing's syndrome is present, correction of the underlying cause will reverse the hypokalemia and alkalosis. $[\text{H}^+]$ loss by the stomach or kidneys can be mitigated by the use of proton pump inhibitors or the discontinuation of diuretics. The second aspect of treatment is to eliminate factors that sustain the inappropriate increase in HCO_3^- reabsorption, such as ECFV contraction or K^+ deficiency. K^+ deficits should always be repaired. Isotonic saline is recommended to reverse the alkalosis when ECFV contraction is present. If associated conditions, such as congestive heart failure, preclude infusion of isotonic saline, renal HCO_3^- loss can be accelerated by administration of acetazolamide (125–250 mg IV), a carbonic anhydrase inhibitor, which is usually effective in patients with adequate renal function. However, acetazolamide triggers urinary K^+

losses and may cause hypokalemia that should be corrected. Dilute hydrochloric acid IV (0.1 N HCl) has been advocated in extreme cases of metabolic alkalosis but causes hemolysis and must be delivered slowly in a central vein. This preparation is not available generally and must be prepared in the pharmacy. Because serious errors or harm may occur, its use is not advised. Therapy in Liddle's syndrome should include a potassium-sparing diuretic (amiloride or triamterene) to inhibit ENaC and correct both the hypertension and the hypokalemia.

RESPIRATORY ACIDOSIS

Respiratory acidosis occurs as a result of severe pulmonary disease, respiratory muscle fatigue, or abnormalities in ventilatory control and is recognized by an increase in Paco_2 and decrease in pH (**Table 55-7**). In acute respiratory acidosis, there is a compensatory elevation in HCO_3^- (due to cellular buffering mechanisms) that increases 1 mmol/L for every 10-mmHg increase in Paco_2 . In chronic respiratory acidosis (>24 h), renal adaptation increases the $[\text{HCO}_3^-]$ by 4 mmol/L for every 10-mmHg increase in Paco_2 . The serum HCO_3^- usually does not increase above 38 mmol/L.

The clinical features vary according to the severity and duration of the respiratory acidosis, the underlying disease, and whether there is accompanying hypoxemia. A rapid increase in Paco_2 (acute hypercapnia) may cause anxiety, dyspnea, confusion, psychosis, and hallucinations and may progress to coma. However, chronic hypercapnia may cause sleep disorders; loss of memory; daytime somnolence; personality changes; impairment of coordination; and motor disturbances such as tremor, myoclonic jerks, and asterixis. Headaches and other signs that mimic raised intracranial pressure, such as papilledema, abnormal reflexes, and focal muscle weakness, are also seen.

Depression of the respiratory center by a variety of drugs, injury, or disease can produce respiratory acidosis. This may occur acutely with general anesthetics, sedatives, and head trauma or chronically with sedatives, alcohol, intracranial tumors, and the syndromes of sleep-disordered breathing including the primary alveolar and obesity-hypoventilation syndromes (**Chaps. 296 and 297**). Abnormalities or disease in the motor neurons, neuromuscular junction, and skeletal muscle can cause hypoventilation via respiratory muscle fatigue. Mechanical ventilation, when not properly adjusted, may result in respiratory acidosis, particularly if CO_2 production suddenly rises (because of fever, agitation, sepsis, or overfeeding) or alveolar ventilation decreases because of worsening pulmonary function. High levels of positive end-expiratory pressure in the presence of reduced cardiac output may cause hypercapnia as a result of large increases in alveolar dead space (**Chap. 285**). Permissive hypercapnia may be used to minimize intrinsic positive end-expiratory pressure in respiratory distress syndrome, but the consequential respiratory acidosis may require administration of NaHCO_3 to increase the arterial pH to approximately 7.20, but not to the normal value.

Acute hypercapnia follows sudden occlusion of the upper airway or generalized bronchospasm as in severe asthma, anaphylaxis, inhalational burn, or toxin injury. Chronic hypercapnia and respiratory acidosis occur in end-stage obstructive lung disease. Restrictive disorders involving both the chest wall and the lungs can cause respiratory acidosis because the high metabolic cost of respiration causes ventilatory muscle fatigue. Advanced stages of intrapulmonary and extrapulmonary restrictive defects present as chronic respiratory acidosis.

The diagnosis of respiratory acidosis requires the measurement of Paco_2 and arterial pH. A detailed history and physical examination often indicate the cause. Pulmonary function studies (**Chap. 285**), including spirometry, diffusion capacity for carbon monoxide, lung volumes, and arterial Paco_2 and O_2 saturation, usually make it possible to determine if respiratory acidosis is secondary to lung disease. The workup for nonpulmonary causes should include a detailed drug history, measurement of hematocrit, and assessment of upper airway, chest wall, pleura, and neuromuscular function.

TABLE 55-7 Respiratory Acid-Base Disorders

- I. Alkalosis
 - A. Central nervous system stimulation
 - 1. Pain
 - 2. Anxiety, psychosis
 - 3. Fever
 - 4. Cerebrovascular accident
 - 5. Meningitis, encephalitis
 - 6. Tumor
 - 7. Trauma
 - B. Hypoxemia or tissue hypoxia
 - 1. High altitude
 - 2. Pneumonia, pulmonary edema
 - 3. Aspiration
 - 4. Severe anemia
 - C. Drugs or hormones
 - 1. Pregnancy, progesterone
 - 2. Salicylates
 - 3. Cardiac failure
 - D. Stimulation of chest receptors
 - 1. Hemothorax
 - 2. Flail chest
 - 3. Cardiac failure
 - 4. Pulmonary embolism
 - E. Miscellaneous
 - 1. Septicemia
 - 2. Hepatic failure
 - 3. Mechanical hyperventilation
 - 4. Heat exposure
 - 5. Recovery from metabolic acidosis
- II. Acidosis
 - A. Central
 - 1. Drugs (anesthetics, morphine, sedatives)
 - 2. Stroke
 - 3. Infection
 - B. Airway
 - 1. Obstruction
 - 2. Asthma
 - C. Parenchyma
 - 1. Emphysema
 - 2. Pneumoconiosis
 - 3. Bronchitis
 - 4. Adult respiratory distress syndrome
 - 5. Barotrauma
 - D. Neuromuscular
 - 1. Poliomyelitis
 - 2. Kyphoscoliosis
 - 3. Myasthenia
 - 4. Muscular dystrophies
 - E. Miscellaneous
 - 1. Obesity
 - 2. Hypoventilation
 - 3. Permissive hypercapnia

TREATMENT

Respiratory Acidosis

The management of respiratory acidosis depends on its severity and rate of onset. Acute respiratory acidosis can be life-threatening, and measures to reverse the underlying cause should be undertaken simultaneously with restoration of adequate alveolar ventilation. This

may necessitate tracheal intubation and assisted mechanical ventilation. Oxygen administration should be titrated carefully in patients with severe obstructive pulmonary disease and chronic CO₂ retention who are breathing spontaneously (*Chap. 292*). When oxygen is used injudiciously, these patients may experience progression of the respiratory acidosis causing severe acidemia. Aggressive and rapid correction of hypercapnia should be avoided, because the falling Paco₂ may provoke the same complications noted with acute respiratory alkalosis (i.e., cardiac arrhythmias, reduced cerebral perfusion, and seizures). The Paco₂ should be lowered gradually in chronic respiratory acidosis, aiming to restore the Paco₂ to baseline levels and to provide sufficient Cl⁻ and K⁺ to enhance the renal excretion of HCO₃⁻.

Chronic respiratory acidosis is frequently difficult to correct, but the primary goal is to institute measures that may improve lung function (*Chap. 292*).

RESPIRATORY ALKALOSIS

Alveolar hyperventilation decreases Paco₂ and increases the HCO₃⁻/Paco₂ ratio, thus increasing pH (Table 55-7). Nonbicarbonate cellular buffers respond by consuming HCO₃⁻. Hypocapnia develops when a sufficiently strong ventilatory stimulus causes CO₂ output in the lungs to exceed its metabolic production by tissues. Plasma pH and [HCO₃⁻] appear to vary proportionately with Paco₂ over a range from 40–15 mmHg. The relationship between arterial [H⁺] concentration and Paco₂ is ~0.7 mmol/L per mmHg (or 0.01 pH unit/mmHg), and that for plasma [HCO₃⁻] is 0.2 mmol/L per mmHg. Hypocapnia sustained for >2–6 h is further compensated by a decrease in renal ammonium and titratable acid excretion and a reduction in filtered HCO₃⁻ reabsorption. Full renal adaptation to respiratory alkalosis may take several days and requires normal volume status and renal function. The kidneys appear to respond directly to the lowered Paco₂ rather than to alkalosis per se. In chronic respiratory alkalosis, a 1-mmHg decrease in Paco₂ causes a 0.4-to 0.5-mmol/L drop in [HCO₃⁻] and a 0.3-mmol/L decrease in [H⁺] (or 0.003 increase in pH).

The effects of respiratory alkalosis vary according to duration and severity but are primarily those of the underlying disease. Reduced cerebral blood flow as a consequence of a rapid decline in Paco₂ may cause dizziness, mental confusion, and seizures, even in the absence of hypoxemia. The cardiovascular effects of acute hypocapnia in the conscious human are generally minimal, but in the anesthetized or mechanically ventilated patient, cardiac output and blood pressure may fall because of the depressant effects of anesthesia and positive-pressure ventilation on heart rate, systemic resistance, and venous return. Cardiac arrhythmias may occur in patients with heart disease as a result of changes in oxygen unloading by blood from a left shift in the hemoglobin-oxygen dissociation curve (Bohr effect). Acute respiratory alkalosis causes intracellular shifts of Na⁺, K⁺, and PO₄²⁻ and reduces free [Ca²⁺] by increasing the protein-bound fraction. Hypocapnia-induced hypokalemia is usually minor.

Chronic respiratory alkalosis is the most common acid-base disturbance in critically ill patients and, when severe, portends a poor prognosis. Many cardiopulmonary disorders manifest respiratory alkalosis in their early to intermediate stages, and the finding of normocapnia and hypoxemia in a patient with hyperventilation may herald the onset of rapid respiratory failure and should prompt an assessment to determine if the patient is becoming fatigued. Respiratory alkalosis is common during mechanical ventilation.

The hyperventilation syndrome may be disabling. Paresthesia; circumoral numbness; chest wall tightness or pain; dizziness; inability to take an adequate breath; and, rarely, tetany may be sufficiently stressful to perpetuate the disorder. Arterial blood-gas analysis demonstrates an acute or chronic respiratory alkalosis, often with hypocapnia in the range of 15–30 mmHg and no hypoxemia. CNS diseases or injury can produce several patterns of hyperventilation and sustained Paco₂

levels of 20–30 mmHg. Hyperthyroidism, high caloric loads, and exercise raise the basal metabolic rate, but ventilation usually rises in proportion so that arterial blood gases are unchanged and respiratory alkalosis does not develop. Salicylates are the most common cause of drug-induced respiratory alkalosis because of direct stimulation of the medullary chemoreceptor (*Chap. 458*). In addition, the methylxanthines, theophylline, and aminophylline stimulate ventilation and increase the ventilatory response to CO₂. Progesterone increases ventilation and lowers arterial Paco₂ by as much as 5–10 mmHg. Therefore, chronic respiratory alkalosis is a common feature of pregnancy. Respiratory alkalosis is also prominent in liver failure, and the severity correlates with the degree of hepatic insufficiency. Respiratory alkalosis is often an early finding in gram-negative septicemia, before fever, hypoxemia, or hypotension develops.

The diagnosis of respiratory alkalosis depends on measurement of arterial pH and Paco₂. The plasma [K⁺] is often reduced and the [Cl⁻] increased. In the acute phase, respiratory alkalosis is not associated with increased renal HCO₃⁻ excretion, but within hours, net acid excretion is reduced. In general, the HCO₃⁻ concentration falls by 2.0 mmol/L for each 10-mmHg decrease in Paco₂. Chronic respiratory alkalosis occurs when hypocapnia persists for greater than 3–5 days. The decline in Paco₂ reduces the serum [HCO₃⁻] by 4.0–5 mmol/L for each 10-mmHg decrease in Paco₂. It is unusual to observe a plasma HCO₃⁻ <12 mmol/L as a result of a pure respiratory alkalosis. The compensatory reduction in plasma [HCO₃⁻] is so effective in chronic respiratory alkalosis that the pH may not decline significantly from the normal value. Therefore, chronic respiratory alkalosis is the only acid-base disorder for which compensation can return the pH to the normal value.

When the diagnosis of respiratory alkalosis is made, its cause should be investigated. The diagnosis of hyperventilation syndrome is made by exclusion. In difficult cases, it may be important to rule out other conditions such as pulmonary embolism, coronary artery disease, and hyperthyroidism.

TREATMENT

Respiratory Alkalosis

The management of respiratory alkalosis is directed toward alleviation of the underlying disorder. If respiratory alkalosis complicates ventilator management, changes in dead space and tidal volume can minimize the hypocapnia. Patients with the hyperventilation syndrome may benefit from reassurance, rebreathing from a paper bag during symptomatic attacks, and attention to underlying psychological stress. Antidepressants and sedatives are not recommended. β-Adrenergic blockers may ameliorate peripheral manifestations of the hyperadrenergic state.

REFERENCES

- BEREND K et al: Physiological approach to assessment of acid-base disturbances. *N Engl J Med* 371:1434, 2014.
- DUBOSE TD: Etiologic causes of metabolic acidosis II: The normal anion gap acidosis. In *Metabolic Acidosis*. Wesson DE (ed). New York, Springer, 2016, pp. 27–38.
- HAMM LL, DUBOSE TD: Disorders of acid-base balance. In *Brenner and Rector's The Kidney*, 11th ed. Yu A et al (eds). Philadelphia, Elsevier, 2020, pp 496–536.
- KRAUT JA, MADIAS NE: Metabolic acidosis of CKD: An update. *Am J Kidney Dis* 67:307, 2016.
- KRAUT JA, MADIAS NE: Re-evaluation of the normal range of serum total CO₂ concentration. *Clin J Am Soc Nephrol* 13:343, 2018.
- PALMER BF, CLEGG DJ: Electrolyte and acid-base disturbances in patients with diabetes mellitus. *N Engl J Med* 373:548, 2015.
- WESSON DE et al: Mechanisms of metabolic acidosis-induced kidney injury in chronic kidney disease. *J Am Soc Nephrol* 31:469, 2020.

Section 8 Alterations in the Skin

56

Approach to the Patient with a Skin Disorder

Kim B. Yancey, Thomas J. Lawley

The challenge of examining the skin lies in distinguishing normal from abnormal findings, distinguishing significant findings from trivial ones, and integrating pertinent signs and symptoms into an appropriate differential diagnosis. The fact that the largest organ in the body is visible is both an advantage and a disadvantage to those who examine it. It is advantageous because no special instrumentation is necessary and because the skin can be biopsied with little morbidity. However, the casual observer can be misled by a variety of stimuli and overlook important, subtle signs of skin or systemic disease. For instance, the sometimes minor differences in color and shape that distinguish a melanoma (Fig. 56-1) from a benign nevomelanocytic nevus (Fig. 56-2) can be difficult to recognize. A variety of descriptive terms have been developed that characterize cutaneous lesions (Tables 56-1, 56-2, and 56-3; Fig. 56-3), thereby aiding in their interpretation and in the formulation of a differential diagnosis (Table 56-4). For example, the finding of scaling papules, which are present in psoriasis or atopic dermatitis, places the patient in a different diagnostic category than would hemorrhagic papules, which may indicate vasculitis or sepsis (Figs. 56-4 and 56-5, respectively). It is also important to differentiate primary from secondary skin lesions. If the examiner focuses on linear erosions overlying an area of erythema and scaling, he or she may incorrectly assume that the erosion is the primary lesion and that the redness and scale are secondary, whereas the correct interpretation would be that the patient has a pruritic eczematous dermatitis with erosions caused by scratching.

APPROACH TO THE PATIENT

Skin Disorder

In examining the skin, it is usually advisable to assess the patient before taking an extensive history. This approach ensures that the entire cutaneous surface will be evaluated, and objective findings can be integrated with relevant historical data. Four basic features of a skin problem must be noted and considered during a physical examination: the *distribution* of the eruption, the *types* of primary



FIGURE 56-1 Superficial spreading melanoma. This is the most common type of melanoma. Such lesions usually demonstrate asymmetry, border irregularity, color variegation (black, blue, brown, pink, and white), a diameter >6 mm, and a history of change (e.g., an increase in size or development of associated symptoms such as pruritus or pain).



FIGURE 56-2 Nevomelanocytic nevus. Nevi are benign proliferations of nevomelanocytes characterized by regularly shaped hyperpigmented macules or papules of a uniform color.

TABLE 56-1 Description of Primary Skin Lesions

Macule: A flat, colored lesion, <2 cm in diameter, not raised above the surface of the surrounding skin. A “freckle,” or ephelid, is a prototypical pigmented macule.

Patch: A large (>2 cm) flat lesion with a color different from the surrounding skin. This differs from a macule only in size.

Papule: A small, solid lesion, <0.5 cm in diameter, raised above the surface of the surrounding skin and thus palpable (e.g., a closed comedone, or whitehead, in acne).

Nodule: A larger (0.5–5.0 cm), firm lesion raised above the surface of the surrounding skin. This differs from a papule only in size (e.g., a large dermal nevomelanocytic nevus).

Tumor: A solid, raised growth >5 cm in diameter.

Plaque: A large (>1 cm), flat-topped, raised lesion; edges may either be distinct (e.g., in psoriasis) or gradually blend with surrounding skin (e.g., in eczematous dermatitis).

Vesicle: A small, fluid-filled lesion, <0.5 cm in diameter, raised above the plane of surrounding skin. Fluid is often visible, and the lesions are translucent (e.g., vesicles in allergic contact dermatitis caused by *Toxicodendron* [poison ivy]).

Pustule: A vesicle filled with leukocytes. Note: The presence of pustules does not necessarily signify the existence of an infection.

Bulla: A fluid-filled, raised, often translucent lesion >0.5 cm in diameter.

Wheal: A raised, erythematous, edematous papule or plaque, usually representing short-lived vasodilation and vasopermeability.

Telangiectasia: A dilated, superficial blood vessel.

TABLE 56-2 Description of Secondary Skin Lesions

Lichenification: A distinctive thickening of the skin that is characterized by accentuated skinfold markings.

Scale: Excessive accumulation of stratum corneum.

Crust: Dried exudate of body fluids that may be either yellow (i.e., serous crust) or red (i.e., hemorrhagic crust).

Erosion: Loss of epidermis without an associated loss of dermis.

Ulcer: Loss of epidermis and at least a portion of the underlying dermis.

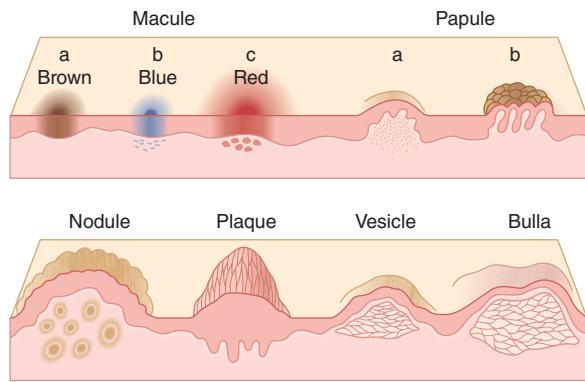
Excoriation: Linear, angular erosions that may be covered by crust and are caused by scratching.

Atrophy: An acquired loss of substance. In the skin, this may appear as a depression with intact epidermis (i.e., loss of dermal or subcutaneous tissue) or as sites of shiny, delicate, wrinkled lesions (i.e., epidermal atrophy).

Scar: A change in the skin secondary to trauma or inflammation. Sites may be erythematous, hypopigmented, or hyperpigmented depending on their age or character. Sites on hair-bearing areas may be characterized by destruction of hair follicles.

TABLE 56-3 Common Dermatologic Terms

Alopecia: Hair loss, partial or complete.
Annular: Ring-shaped.
Cyst: A soft, raised, encapsulated lesion filled with semisolid or liquid contents.
Herpetiform: In a grouped configuration.
Lichenoid eruption: Violaceous to purple, polygonal lesions that resemble those seen in lichen planus.
Milia: Small, firm, white papules filled with keratin.
Morbilliform rash: Generalized, small erythematous macules and/or papules that resemble lesions seen in measles.
Nummular: Coin-shaped.
Poikiloderma: Skin that displays variegated pigmentation, atrophy, and telangiectasias.
Polycyclic lesions: A configuration of skin lesions formed from coalescing rings or incomplete rings.
Pruritus: A sensation that elicits the desire to scratch. Pruritus is often the predominant symptom of inflammatory skin diseases (e.g., atopic dermatitis, allergic contact dermatitis); it is also commonly associated with xerosis and aged skin. Systemic conditions that can be associated with pruritus include chronic renal disease, cholestasis, pregnancy, malignancy, thyroid disease, polycythemia vera, and delusions of parasitosis.

**FIGURE 56-3 A schematic representation of several common primary skin lesions (see Table 56-1).****TABLE 56-4 Selected Common Dermatologic Conditions**

DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY	DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY
Acne vulgaris	Face, upper back, chest	Open and closed comedones, erythematous papules, pustules, cysts	Seborrheic keratosis	Trunk, face, extremities	Brown plaques with adherent, greasy scale, "stuck on" appearance
Rosacea	Blush area of cheeks, nose, forehead, chin	Erythema, telangiectasias, papules, pustules	Folliculitis Impetigo	Any hair-bearing area Anywhere	Follicular pustules Papules, vesicles, pustules, often with honey-colored crusts
Seborrheic dermatitis	Scalp, eyebrows, perinasal areas	Erythema with greasy yellow-brown scale	Herpes simplex	Lips, genitalia	Grouped vesicles progressing to crusted erosions
Atopic dermatitis	Antecubital and popliteal fossae; may be widespread	Patches and plaques of erythema, scaling, and lichenification; pruritus	Herpes zoster	Dermatomal, usually trunk but may be anywhere	Vesicles limited to a dermatome (often painful)
Stasis dermatitis	Ankles, lower legs over medial malleoli	Patches of erythema and scaling on background of hyperpigmentation associated with signs of venous insufficiency	Varicella	Face, trunk, relative sparing of extremities	Lesions arise in crops and quickly progress from erythematous macules, to papules, to vesicles, to pustules, to crusted sites
Dyshidrotic eczema	Palms, soles, sides of fingers, and toes	Deep vesicles	Pityriasis rosea	Trunk (Christmas tree pattern); herald patch followed by multiple smaller lesions	Symmetric erythematous papules and plaques with a collarette of scale
Allergic contact dermatitis	Anywhere	Localized erythema, vesicles, scale, and pruritus (e.g., fingers, earlobes—nickel; dorsal aspect of foot—shoe; exposed surfaces—poison ivy)	Tinea versicolor	Chest, back, abdomen, proximal extremities	Scaly hyper- or hypopigmented macules
Psoriasis	Elbows, knees, scalp, lower back, fingernails (may be generalized)	Papules and plaques covered with silvery scale; nails have pits	Candidiasis	Groin, beneath breasts, vagina, oral cavity	Erythematous macerated areas with satellite pustules; white, friable patches on mucous membranes
Lichen planus	Wrists, ankles, mouth (may be widespread)	Violaceous flat-topped papules and plaques	Dermatophytosis	Feet, groin, beard, or scalp	Varies with site (e.g., tinea corporis—scaly annular plaque)
Keratosis pilaris	Extensor surfaces of arms and thighs, buttocks	Keratotic follicular papules with surrounding erythema	Scabies	Groin, axillae, between fingers and toes, beneath breasts	Excoriated papules, burrows, pruritus
Melasma	Forehead, cheeks, temples, upper lip	Tan to brown patches	Insect bites	Anywhere	Erythematous papules with central puncta
Vitiligo	Periorificial, trunk, extensor surfaces of extremities, flexor wrists, axillae	Chalk-white macules	Cherry angioma Keloid Dermatofibroma	Trunk Anywhere (site of previous injury) Anywhere	Red, blood-filled papules Firm tumor, pink, purple, or brown Firm red to brown nodule that shows dimpling of overlying skin with lateral compression

(Continued)

TABLE 56-4 Selected Common Dermatologic Conditions (Continued)

DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY	DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY
Actinic keratosis	Sun-exposed areas	Skin-colored or red-brown macule or papule with dry, rough, adherent scale	Acrochordons (skin tags)	Groin, axilla, neck	Fleshy papules
Basal cell carcinoma	Face	Papule with pearly, telangiectatic border on sun-damaged skin	Urticaria	Anywhere	Wheals, sometimes with surrounding flare; pruritus
Squamous cell carcinoma	Face, especially lower lip, ears	Indurated and possibly hyperkeratotic lesions often showing ulceration and/or crusting	Transient acantholytic dermatosis Xerosis	Trunk, especially anterior chest Extensor extremities, especially legs	Erythematous papules Dry, erythematous, scaling patches; pruritus

and secondary lesions, the *shape* of individual lesions, and the *arrangement* of the lesions. An ideal skin examination includes evaluation of the skin, hair, and nails as well as the mucous membranes of the mouth, eyes, nose, nasopharynx, and anogenital region. In the initial examination, it is important that the patient be disrobed as completely as possible to minimize chances of missing important individual skin lesions and permit accurate assessment of the distribution of the eruption. The patient should first be viewed from a distance of about 1.5–2 m (4–6 ft) so that the general character of the skin and the distribution of lesions can be evaluated. Indeed, the distribution of lesions often correlates highly with diagnosis (Fig. 56-6). For example, a hospitalized patient with a generalized erythematous exanthem is more likely to have a drug eruption than is a patient with a similar rash limited to the sun-exposed portions of the face. Once the distribution of the lesions has been established, the nature of the primary lesion must be determined. Thus, when lesions are distributed on elbows, knees, and scalp, the most likely possibility based solely on distribution is psoriasis or dermatitis herpetiformis (Figs. 56-7 and 56-8, respectively). The primary lesion in psoriasis is a scaly papule that soon forms erythematous plaques covered with a white scale, whereas that of dermatitis herpetiformis is an urticarial papule that quickly becomes a small vesicle. In this manner, identification of the primary lesion directs the examiner toward the proper diagnosis. Secondary changes in skin can also be quite helpful. For example, scale represents excessive epidermis, while crust is the result of a discontinuous epithelial cell layer. Palpation of skin lesions can yield insight into the character of an eruption. Thus, red papules on the lower extremities that blanch

with pressure can be a manifestation of many different diseases, but hemorrhagic red papules that do not blanch with pressure indicate palpable purpura characteristic of necrotizing vasculitis (Fig. 56-4).

The shape of lesions is also an important feature. Flat, round, erythematous papules and plaques are common in many cutaneous diseases. However, target-shaped lesions that consist in part of erythematous plaques are specific for erythema multiforme (Fig. 56-9). Likewise, the arrangement of individual lesions is important. Erythematous papules and vesicles can occur in many conditions, but their arrangement in a specific linear array suggests an external etiology such as allergic contact dermatitis (Fig. 56-10) or primary irritant dermatitis. In contrast, lesions with a generalized arrangement are common and suggest a systemic etiology.

As in other branches of medicine, a complete history should be obtained to emphasize the following features:

1. Evolution of lesions
 - a. Site of onset
 - b. Manner in which the eruption progressed or spread
 - c. Duration
 - d. Periods of resolution or improvement in chronic eruptions
2. Symptoms associated with the eruption
 - a. Itching, burning, pain, numbness
 - b. What, if anything, has relieved symptoms
 - c. Time of day when symptoms are most severe
3. Current or recent medications (prescribed as well as over-the-counter)
4. Associated systemic symptoms (e.g., malaise, fever, arthralgias)
5. Ongoing or previous illnesses
6. History of allergies
7. Presence of photosensitivity
8. Review of systems
9. Family history (particularly relevant for patients with melanoma, atopy, psoriasis, or acne)
10. Social, sexual, or travel history



FIGURE 56-4 Necrotizing vasculitis. Palpable purpuric papules on the lower legs are seen in this patient with cutaneous small-vessel vasculitis. (Courtesy of Robert Swerlick, MD; with permission.)



FIGURE 56-5 Meningococcemia. An example of fulminant meningococcemia with extensive angular purpuric patches. (Courtesy of Stephen E. Gellis, MD; with permission.)

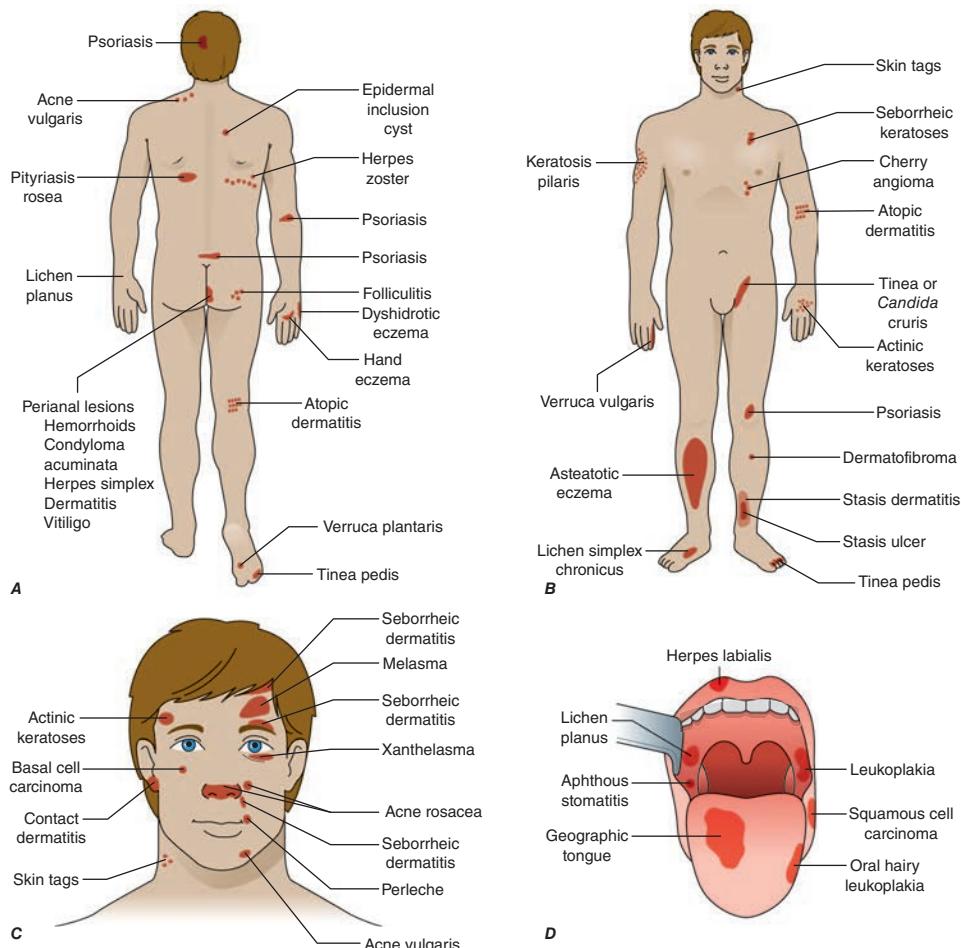


FIGURE 56-6 Distribution of some common dermatologic diseases and lesions.

■ DIAGNOSTIC TECHNIQUES

Many skin diseases can be diagnosed on the basis of gross clinical appearance, but sometimes relatively simple diagnostic procedures can yield valuable information. In most instances, they can be performed at the bedside with a minimum of equipment.

Skin Biopsy A skin biopsy is a straightforward minor surgical procedure; however, it is important to biopsy a lesion that is most likely to yield diagnostic findings. This decision may require expertise in skin diseases and knowledge of superficial anatomic structures in selected areas of the body. In this procedure, a small area of skin is anesthetized with 1% lidocaine with or without epinephrine. The skin lesion in

question can be excised or saucerized with a scalpel or removed by punch biopsy. In the latter technique, a punch is pressed against the surface of the skin and rotated with downward pressure until it penetrates to the subcutaneous tissue. The circular biopsy is then lifted with forceps, and the bottom is cut with iris scissors. Biopsy sites may or may not need suture closure, depending on size and location.

KOH Preparation A potassium hydroxide (KOH) preparation is performed on scaling skin lesions where a fungal infection is suspected. The edge of such a lesion is scraped gently with a no. 15 scalpel blade. The removed scale is collected on a glass microscope slide, treated with 1 or 2 drops of a solution of 10–20% KOH, and placement of a cover



FIGURE 56-7 Psoriasis. This papulosquamous skin disease is characterized by small and large erythematous papules and plaques with overlying adherent silvery scale.



FIGURE 56-8 Dermatitis herpetiformis. This disorder typically displays pruritic, grouped papulovesicles on elbows, knees, buttocks, and posterior scalp. Vesicles are often exorciated due to associated pruritus.



FIGURE 56-9 Erythema multiforme. This eruption is characterized by multiple erythematous plaques with a target or iris morphology. It usually represents a hypersensitivity reaction to drugs (e.g., sulfonamides) or infections (e.g., HSV). (Courtesy of the Yale Resident's Slide Collection; with permission.)

slip. KOH dissolves keratin and allows easier visualization of fungal elements. Brief heating of the slide accelerates dissolution of keratin. When the preparation is viewed under the microscope, the refractile hyphae are seen more easily when the light intensity is reduced and the condenser is lowered. This technique can be used to identify hyphae in dermatophyte infections, pseudothelia and budding yeasts in *Candida* infections, and “spaghetti and meatballs” yeast forms in *Tinea versicolor*. The same sampling technique can be used to obtain scale for culture of selected pathogenic organisms.

Tzanck Smear A Tzanck smear is a cytologic technique most often used in the diagnosis of herpesvirus infections (herpes simplex virus [HSV] or varicella-zoster virus [VZV]) (see Figs. 193-1 and 193-3). An early vesicle, not a pustule or crusted lesion, is unroofed, and the base of the lesion is scraped gently with a scalpel blade. The material is



FIGURE 56-11 Urticaria. Discrete and confluent, edematous, erythematous papules and plaques are characteristic of this whealing eruption.

placed on a glass slide, air-dried, and stained with Giemsa or Wright's stain. Multinucleated epithelial giant cells suggest the presence of HSV or VZV; culture, immunofluorescence microscopy, or genetic testing must be performed to identify the specific virus.

Diascopy Diascopy is designed to assess whether a skin lesion will blanch with pressure as, for example, in determining whether a red lesion is hemorrhagic or simply blood-filled. Urticaria (Fig. 56-11) will blanch with pressure, whereas a purpuric lesion caused by necrotizing vasculitis (Fig. 56-4) will not. Diascopy is performed by pressing a microscope slide or magnifying lens against a lesion and noting the amount of blanching that occurs. Granulomas often have an opaque to transparent, brown-pink “apple jelly” appearance on diascopy.

Dermoscopy Dermoscopy is a noninvasive method of examining the skin surface that uses a high-quality magnifying lens and a specialized light source (i.e., a dermatoscope). Dermoscopy identifies skin structures, colors, and patterns that are not visible to the naked eye. It is particularly useful in the evaluation of pigmented skin lesions.

Wood's Light A Wood's lamp generates 360-nm ultraviolet (“black”) light that can be used to aid the evaluation of certain skin disorders. For example, a Wood's lamp will cause erythrasma (a superficial, intertriginous infection caused by *Corynebacterium minutissimum*) to show a characteristic coral pink color, and wounds colonized by *Pseudomonas* will appear pale blue. Tinea capitis caused by certain dermatophytes (e.g., *Microsporum canis* or *M. audouinii*) exhibits a yellow fluorescence. Pigmented lesions of the epidermis such as freckles are accentuated, while dermal pigment such as postinflammatory hyperpigmentation fades under a Wood's light. Vitiligo (Fig. 56-12)



FIGURE 56-10 Allergic contact dermatitis (ACD). **A.** An example of ACD in its acute phase, with sharply demarcated, weeping, eczematous plaques in a perioral distribution. **B.** ACD in its chronic phase, with an erythematous, lichenified, weeping plaque on skin chronically exposed to nickel in a metal snap. (**B**, Courtesy of Robert Swerlick, MD; with permission.)



FIGURE 56-12 Vitiligo. Characteristic lesions display an acral distribution and striking depigmentation as a result of loss of melanocytes.

appears totally white under a Wood's lamp, and previously unsuspected areas of involvement often become apparent. A Wood's lamp may also aid in the demonstration of tinea versicolor, detection of sites of depigmentation within and/or surrounding melanomas, and recognition of ash leaf spots in patients with tuberous sclerosis.

Patch Tests Patch testing is designed to document sensitivity to a specific antigen. In this procedure, a battery of suspected allergens is applied to the patient's back under occlusive dressings and allowed to remain in contact with the skin for 48 h. The dressings are removed, and the area is examined for evidence of delayed hypersensitivity reactions (e.g., erythema, edema, or papulovesicles). This test is best performed by physicians with special expertise in patch testing and is often helpful in the evaluation of patients with chronic dermatitis.

FURTHER READING

- BOLOGNIA JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.
 JAMES WD: *Andrews' Diseases of the Skin: Clinical Dermatology*, 13th ed. Philadelphia, Elsevier, 2019.
 KANG S et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 9th ed. New York, McGraw-Hill, 2019.

57

Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders

Leslie P. Lawley, Justin T. Cheeley,
 Robert A. Swerlick

TABLE 57-1 Clinical Features of Atopic Dermatitis

1. Pruritus and scratching
2. Course marked by exacerbations and remissions
3. Lesions typical of eczematous dermatitis
4. Personal or family history of atopy (asthma, allergic rhinitis, food allergies, or eczema)
5. Clinical course lasting >6 weeks
6. Lichenification of skin
7. Presence of dry skin

The clinical presentation often varies with age. Half of patients with AD present within the first year of life, and 80% present by 5 years of age. About 80% ultimately coexpress allergic rhinitis or asthma. The infantile pattern is characterized by weeping inflammatory patches and crusted plaques on the face, neck, and extensor surfaces. The childhood and adolescent patterns are typified by dermatitis of flexural skin, particularly in the antecubital and popliteal fossae (Fig. 57-1). AD may resolve spontaneously, but approximately 40% of all individuals affected as children will have dermatitis in adult life. The distribution of lesions in adults may be similar to those seen in childhood; however, adults frequently have localized disease manifesting as lichen simplex chronicus or hand eczema (see below). In patients with localized disease, AD may be suspected because of a typical personal or family history or the presence of cutaneous stigmata of AD such as perioral pallor, an extra fold of skin beneath the lower eyelid (Dennie-Morgan folds), increased palmar skin markings, and an increased incidence of cutaneous infections, particularly with *Staphylococcus aureus*. Regardless of other manifestations, pruritus is a prominent characteristic of AD in all age groups and is exacerbated by dry skin. Many of the cutaneous findings in affected patients, such as lichenification, are secondary to rubbing and scratching.

TREATMENT

Atopic Dermatitis

Therapy for AD should include avoidance of cutaneous irritants, adequate moisturization through the application of emollients, judicious use of topical anti-inflammatory agents, and prompt treatment of secondary infection. Patients should be instructed to bathe no more often than daily, using warm or cool water, and to use only mild bath soap. Immediately after bathing, while the skin is still moist, a topical anti-inflammatory agent in a cream or ointment base should be applied to areas of dermatitis, and all other skin areas should be lubricated with a moisturizer. Approximately 30 g of a topical agent is required to cover the entire body surface of an average adult.



FIGURE 57-1 Atopic dermatitis. Hyperpigmentation, lichenification, and scaling in the antecubital fossae are seen in this patient with atopic dermatitis. (Courtesy of Robert Swerlick, MD.)

ECZEMA AND DERMATITIS

Eczema is a type of dermatitis, and these terms are often used synonymously (e.g., atopic eczema or atopic dermatitis [AD]). Eczema is a reaction pattern that presents with variable clinical findings and the common histologic finding of *spongiosis* (intercellular edema of the epidermis). Eczema is the final common expression for a number of disorders, including those discussed in the following sections. Primary lesions may include erythematous macules, papules, and vesicles, which can coalesce to form patches and plaques. In severe eczema, secondary lesions from infection or excoriation, marked by weeping and crusting, may predominate. In chronic eczematous conditions, *lichenification* (cutaneous hypertrophy and accentuation of normal skin markings) may alter the characteristic appearance of eczema.

ATOPIC DERMATITIS

AD is the cutaneous expression of the atopic state, characterized by a family history of asthma, allergic rhinitis, or eczema. The prevalence of AD is increasing worldwide. Some of its features are shown in Table 57-1.

 The etiology of AD is only partially defined, but there is a clear genetic predisposition. When both parents are affected by AD, >80% of their children manifest the disease. When only one parent is affected, the prevalence drops to slightly >50%. A characteristic defect in AD that contributes to the pathophysiology is an impaired epidermal barrier. In many patients, a mutation in the gene encoding filaggrin, a structural protein in the stratum corneum, is responsible. Patients with AD may display a variety of immunoregulatory abnormalities, including increased IgE synthesis; increased serum IgE levels; and impaired, delayed-type hypersensitivity reactions.

Low- to mid-potency topical glucocorticoids are employed in most treatment regimens for AD. Skin atrophy and the potential for systemic absorption are constant concerns, especially with more potent agents. Low-potency topical glucocorticoids or nonglucocorticoid anti-inflammatory agents should be selected for use on the face and in intertriginous areas to minimize the risk of skin atrophy. Three nonglucocorticoid anti-inflammatory agents approved by the U.S. Food and Drug Administration (FDA) are available for topical use in AD: tacrolimus ointment, pimecrolimus cream, and crisaborole ointment. These agents do not cause skin atrophy, nor do they suppress the hypothalamic-pituitary-adrenal axis. The first two agents are topical calcineurin inhibitors (TCIs), whereas crisaborole is a phosphodiesterase-4 inhibitor. Concerns regarding the potential for lymphomas in patients treated with TCIs have largely been unfounded. Currently, all three agents are more costly than topical glucocorticoids. Barrier-repair products that attempt to restore the impaired epidermal barrier are also nonglucocorticoid agents and are gaining popularity in the treatment of AD.

Secondary infection of eczematous skin may lead to exacerbation of AD. Crusted and weeping skin lesions may be infected with *S. aureus*. When secondary infection is suspected, eczematous lesions should be cultured and patients treated with systemic antibiotics active against *S. aureus*. The initial use of penicillinase-resistant penicillins or cephalosporins is preferable. Dicloxacillin or cephalexin (250 mg qid for 7–10 days) is generally adequate for adults; however, antibiotic selection must be directed by culture results and clinical response. More than 50% of *S. aureus* isolates are now methicillin resistant in some communities. Current recommendations for the treatment of infection with these community-acquired methicillin-resistant *S. aureus* (CA-MRSA) strains in adults include trimethoprim-sulfamethoxazole (one double-strength tablet bid), minocycline (100 mg bid), doxycycline (100 mg bid), or clindamycin (300–450 mg qid). Duration of therapy should be 7–10 days. Inducible resistance may limit clindamycin's usefulness. Such resistance can be detected by the double-disk diffusion test, which should be ordered if the isolate is erythromycin resistant and clindamycin sensitive. As an adjunct, antibacterial washes or dilute sodium hypochlorite baths (0.005% bleach) and intermittent nasal mupirocin may be useful.

Control of pruritus is essential for treatment, as AD often represents “an itch that rashes.” Antihistamines are most often used to control pruritus. Diphenhydramine (25 mg every 4–6 h), hydroxyzine (10–25 mg every 6 h), and doxepin (10–25 mg at bedtime) are useful primarily due to their sedating action. Higher doses of these agents may be required, but sedation can become bothersome. Patients need to be counseled about driving or operating heavy equipment after taking these medications. When used at bedtime, sedating antihistamines may improve the patient's sleep. Although they are effective in urticaria, nonsedating antihistamines and selective H₂ blockers are of little use in controlling the pruritus of AD.

Treatment with systemic glucocorticoids should be limited to severe exacerbations unresponsive to topical therapy. In the patient with chronic AD, therapy with systemic glucocorticoids will generally clear the skin only briefly, and cessation of the systemic therapy will invariably be accompanied by a return, if not a worsening, of the dermatitis. For chronic severe AD poorly responsive to standard topical regimens, systemic agents may be considered. Cyclosporine is approved for treatment of severe recalcitrant AD in some European countries. Monitoring of renal function and secondary infections is required. Dupilumab, an interleukin 4 receptor blocker, is FDA approved for use in patients 6 years of age and older and provides more targeted immunomodulation and a better safety profile than cyclosporine. Patients who do not respond to conventional therapies should be considered for patch testing to rule out allergic contact dermatitis (ACD). The role of dietary allergens in AD is controversial, and there is little evidence that they play any role outside of infancy, during which a small percentage of patients with AD may be affected by food allergens.

■ LICHEN SIMPLEX CHRONICUS

Lichen simplex chronicus may represent the end stage of a variety of pruritic and eczematous disorders, including AD. It consists of a circumscribed plaque or plaques of lichenified skin due to chronic scratching or rubbing. Common areas involved include the posterior nuchal region, dorsum of the feet, and ankles. Treatment of lichen simplex chronicus centers on breaking the cycle of chronic itching and scratching. High-potency topical glucocorticoids are helpful in most cases, but, in recalcitrant cases, application of topical glucocorticoids under occlusion or intralesional injection of glucocorticoids may be required.

■ CONTACT DERMATITIS

Contact dermatitis is an inflammatory skin process caused by an exogenous agent or agents that directly or indirectly injure the skin. In **irritant contact dermatitis (ICD)**, this injury is caused by an inherent characteristic of a compound—for example, a concentrated acid or base. Agents that cause **allergic contact dermatitis (ACD)** induce an antigen-specific immune response (e.g., poison ivy dermatitis). The clinical lesions of contact dermatitis may be acute (wet and edematous) or chronic (dry, thickened, and scaly), depending on the persistence of the insult (see Chap. 56, Fig. 56-10).

Irritant Contact Dermatitis ICD is generally well demarcated and often localized to areas of thin skin (eyelids, intertriginous areas) or areas where the irritant was occluded. Lesions may range from minimal skin erythema to areas of marked edema, vesicles, and ulcers. Prior exposure to the offending agent is not necessary, and the reaction develops in minutes to a few hours. Chronic low-grade irritant dermatitis is the most common type of ICD, and the most common area of involvement is the hands (see below). The most common irritants encountered are chronic wet work, soaps, and detergents. Treatment should be directed toward the avoidance of irritants and the use of protective gloves or clothing.

Allergic Contact Dermatitis ACD is a manifestation of delayed-type hypersensitivity mediated by memory T lymphocytes in the skin. Prior exposure to the offending agent is necessary to develop the hypersensitivity reaction, which may take as little as 12 h or as long as 72 h to develop. The most common cause of ACD is exposure to plants, especially to members of the family Anacardiaceae, including the genus *Toxicodendron*. Poison ivy, poison oak, and poison sumac are members of this genus and cause an allergic reaction marked by erythema, vesication, and severe pruritus. The eruption is often linear or angular, corresponding to areas where plants have touched the skin. The sensitizing antigen common to these plants is urushiol, an oleoresin containing the active ingredient pentadecylcatechol. The oleoresin may adhere to skin, clothing, tools, and pets, and contaminated articles may cause dermatitis even after prolonged storage. Blister fluid does not contain urushiol and is not capable of inducing skin eruption in exposed subjects.

TREATMENT

Contact Dermatitis

If contact dermatitis is suspected and an offending agent is identified and removed, the eruption will resolve. Usually, treatment with high-potency topical glucocorticoids is enough to relieve symptoms while the dermatitis runs its course. For patients who require systemic therapy, daily oral prednisone—beginning at 1 mg/kg, but usually ≤60 mg/d—is sufficient. The dose should be tapered over 2–3 weeks, and each daily dose should be taken in the morning with food.

Identification of a contact allergen can be a difficult and time-consuming task. ACD should be suspected in patients with dermatitis unresponsive to conventional therapy or with an unusual and patterned distribution. Patients should be questioned carefully regarding occupational exposures and topical medications. Common sensitizers include preservatives in topical preparations,



FIGURE 57-2 Dyshidrotic eczema. This example is characterized by deep-seated vesicles and scaling on palms and lateral fingers, and the disease is often associated with an atopic diathesis.

nickel sulfate, potassium dichromate, thimerosal, neomycin sulfate, fragrances, formaldehyde, and rubber-curing agents. Patch testing is helpful in identifying these agents but should not be attempted when patients have widespread active dermatitis or are taking systemic glucocorticoids.

HAND ECZEMA

Hand eczema is a very common, chronic skin disorder in which both exogenous and endogenous factors play important roles. It may be associated with other cutaneous disorders such as AD, and contact with various agents may be involved. Hand eczema represents a large proportion of cases of occupation-associated skin disease. Chronic, excessive exposure to water and detergents, harsh chemicals, or allergens may initiate or aggravate this disorder. It may present with dryness and cracking of the skin of the hands as well as with variable amounts of erythema and edema. Often, the dermatitis will begin under rings, where water and irritants are trapped. *Dyshidrotic* eczema, a variant of hand eczema, presents with multiple, intensely pruritic, small papules and vesicles on the thenar and hypothenar eminences and the sides of the fingers (Fig. 57-2). Lesions tend to occur in crops that slowly form crusts and then heal.

The evaluation of a patient with hand eczema should include an assessment of potential occupation-associated exposures. The history should be directed to identifying possible irritant or allergen exposures.

TREATMENT

Hand Eczema

Therapy for hand eczema is directed toward avoidance of irritants, identification of possible contact allergens, treatment of coexistent infection, and application of topical glucocorticoids. Whenever possible, the hands should be protected by gloves, preferably vinyl. The use of rubber gloves (latex) to protect dermatitic skin is sometimes associated with the development of hypersensitivity reactions to components of the gloves, which could be either a type I hypersensitivity reaction to the latex (manifested by the development of hives, itching, angioedema, and possibly anaphylaxis within minutes to hours of exposure) or a type IV hypersensitivity reaction to rubber accelerators (with worsening of eczematous eruptions days after exposure). Patients can be treated with cool moist compresses followed by application of a mid- to high-potency topical glucocorticoid in a cream or ointment base. As in AD, treatment of secondary infection is essential for good control. In addition, patients with hand eczema should be examined for dermatophyte infection by potassium hydroxide (KOH) preparation and culture (see below).

NUMMULAR ECZEMA

Nummular eczema is characterized by circular or oval “coinlike” lesions, beginning as small edematous papules that become crusted and scaly. The etiology of nummular eczema is unknown, but dry skin is a contributing factor. Common locations are the trunk or the extensor surfaces of the extremities, particularly on the pretibial areas or dorsum of the hands. Nummular eczema occurs more frequently in men and is most common in middle age. The treatment of nummular eczema is similar to that for AD.

ASTEATOTIC ECZEMA

Asteatotic eczema, also known as *xerotic eczema* or “winter itch,” is a mildly inflammatory dermatitis that develops in areas of extremely dry skin, especially during the dry winter months. Clinically, there may be considerable overlap with nummular eczema. This form of eczema accounts for many physician visits because of the associated pruritus. Fine cracks and scale, with or without erythema, characteristically develop in areas of dry skin, especially on the anterior surfaces of the lower extremities in elderly patients. Asteatotic eczema responds well to topical moisturizers and the avoidance of cutaneous irritants. Over-bathing and the use of harsh soaps exacerbate asteatotic eczema.

STASIS DERMATITIS AND STASIS ULCERATION

Stasis dermatitis develops on the lower extremities secondary to venous incompetence and chronic edema. Patients may give a history of deep venous thrombosis and may have evidence of vein removal or varicose veins. Early findings in stasis dermatitis consist of mild erythema and scaling associated with pruritus. The typical initial site of involvement is the medial aspect of the ankle, often over a distended vein (Fig. 57-3).

Stasis dermatitis may become acutely inflamed, with crusting and exudate. In this state, it is easily confused with cellulitis. Of note, symmetrical and bilateral involvement is more likely stasis dermatitis, whereas unilateral involvement may represent cellulitis. Chronic stasis dermatitis is often associated with dermal fibrosis that is recognized clinically as brawny edema of the skin. As the disorder progresses, the dermatitis becomes progressively pigmented due to chronic erythrocyte extravasation leading to cutaneous hemosiderin deposition. Stasis dermatitis may be complicated by secondary infection and contact dermatitis. Severe stasis dermatitis may precede the development of stasis ulcers.

TREATMENT

Stasis Dermatitis and Stasis Ulceration

Patients with stasis dermatitis and stasis ulceration benefit greatly from leg elevation and the routine use of compression stockings with a gradient of at least 30–40 mmHg. Stockings providing less



FIGURE 57-3 Stasis dermatitis. An example of stasis dermatitis showing erythematous, scaly, and oozing patches over the lower leg. Several stasis ulcers are also seen in this patient.



FIGURE 57-4 Seborrheic dermatitis. Central facial erythema with overlying greasy, yellowish scale is seen in this patient. (Courtesy of Jean Bolognia, MD; with permission.)

compression, such as antiembolism hose, are poor substitutes. Use of emollients and/or mid-potency topical glucocorticoids and avoidance of irritants are also helpful in treating stasis dermatitis. Protection of the legs from injury, including scratching, and control of chronic edema are essential to prevent ulcers. Diuretics may be required to adequately control chronic edema.

Stasis ulcers are difficult to treat, and resolution is slow. It is extremely important to elevate the affected limb as much as possible. The ulcer should be kept clear of necrotic material by gentle debridement and covered with a semipermeable dressing and a compression dressing or compression stocking. Glucocorticoids should not be applied to ulcers, because they may retard healing; however, they may be applied to the surrounding skin to control itching, scratching, and additional trauma. Superficial bacterial cultures of chronic stasis ulcers often yield polymicrobial colonizers and are of little utility in determination of secondary infection. Care must be taken to exclude treatable causes of leg ulcers (hypercoagulation, vasculitis, arterial insufficiency) before beginning the chronic management outlined above.

■ SEBORRHEIC DERMATITIS

Seborrheic dermatitis is a common, chronic disorder characterized by greasy scales overlying erythematous patches or plaques. Induration and scale are generally less prominent than in psoriasis, but clinical overlap exists between these diseases ("sebopsoriasis"). The most common location is in the scalp, where it may be recognized as severe dandruff. On the face, seborrheic dermatitis affects the eyebrows, eyelids, glabella, and nasolabial folds (Fig. 57-4). Scaling of the external

auditory canal is common in seborrheic dermatitis. In addition, the postauricular areas often become macerated and tender. Seborrheic dermatitis may also develop in the central chest, axilla, groin, submammary folds, and gluteal cleft. Rarely, it may cause widespread generalized dermatitis. Pruritus is variable.

Seborrheic dermatitis may be evident within the first few weeks of life, and within this context, it typically occurs in the scalp ("cradle cap"), face, or groin. It is rarely seen in children beyond infancy but becomes evident again during adolescent and adult life. Although it is frequently seen in patients with Parkinson's disease, in those who have had cerebrovascular accidents, and in those with HIV infection, the overwhelming majority of individuals with seborrheic dermatitis have no underlying disorder.

TREATMENT

Seborrheic Dermatitis

Treatment with low-potency topical glucocorticoids in conjunction with a topical antifungal agent, such as ketoconazole cream or ciclopirox cream, is often effective. The scalp and beard areas may benefit from antidandruff shampoos, which should be left in place 3–5 min before rinsing. High-potency topical glucocorticoid solutions (betamethasone or clobetasol) are effective for control of severe scalp involvement. High-potency glucocorticoids should not be used on the face because this treatment is often associated with steroid-induced rosacea or atrophy.

PAPULOSQUAMOUS DISORDERS (TABLE 57-2)

■ PSORIASIS

Psoriasis is one of the most common dermatologic diseases, affecting up to 2% of the world's population. It is an immune-mediated disease clinically characterized by erythematous, sharply demarcated papules and rounded plaques covered by silvery micaceous scale. The skin lesions of psoriasis are variably pruritic. Traumatized areas often develop lesions of psoriasis (the Koebner or isomorphic phenomenon). In addition, other external factors may exacerbate psoriasis, including infections, stress, and medications (lithium, beta blockers, and antimalarial drugs).

The most common variety of psoriasis is called *plaque-type*. Patients with plaque-type psoriasis have stable, slowly enlarging plaques, which remain basically unchanged for long periods of time. The most commonly involved areas are the elbows, knees, gluteal cleft, and scalp. Involvement tends to be symmetric. Plaque psoriasis generally develops slowly and runs an indolent course. It rarely remits spontaneously. *Inverse psoriasis* affects the intertriginous regions, including the axilla, groin, submammary region, and navel; it also tends to affect the scalp, palms, and soles. The individual lesions are sharply demarcated plaques (see Chap. 56, Fig. 56-7), but they may be moist and without scale due to their locations.

TABLE 57-2 Papulosquamous Disorders

	CLINICAL FEATURES	OTHER NOTABLE FEATURES	HISTOLOGIC FEATURES
Psoriasis	Sharply demarcated, erythematous plaques with mica-like scale; predominantly on elbows, knees, and scalp; atypical forms may localize to intertriginous areas; eruptive forms may be associated with infection	May be aggravated by certain drugs, infection; severe forms seen in association with HIV	Acanthosis, vascular proliferation
Lichen planus	Purple polygonal papules marked by severe pruritus; lacy white markings, especially associated with mucous membrane lesions	Certain drugs may induce: thiazides, antimalarial drugs	Interface dermatitis
Pityriasis rosea	Rash often preceded by herald patch; oval to round plaques with trailing scale; most often affects trunk; eruption lines up in skinfolds giving a "fir tree-like" appearance; generally spares palms and soles	Variable pruritus; self-limited, resolving in 2–8 weeks; may be imitated by secondary syphilis	Pathologic features often nonspecific
Dermatophytosis	Polymorphous appearance depending on dermatophyte, body site, and host response; sharply defined to ill-demarcated scaly plaques with or without inflammation; may be associated with hair loss	KOH preparation may show branching hyphae; culture helpful	Hyphae and neutrophils in stratum corneum

Abbreviations: HIV, human immunodeficiency virus; KOH, potassium hydroxide.

Guttate psoriasis (eruptive psoriasis) is most common in children and young adults. It develops acutely in individuals without psoriasis or in those with chronic plaque psoriasis. Patients present with many small erythematous, scaling papules, frequently after upper respiratory tract infection with β -hemolytic streptococci. The differential diagnosis should include pityriasis rosea and secondary syphilis.

In *pustular psoriasis*, patients may have disease localized to the palms and soles, or the disease may be generalized. Regardless of the extent of disease, the skin is erythematous, with pustules and variable scale. Localized to the palms and soles, it is easily confused with dishydrrotic eczema. When it is generalized, episodes are characterized by fever (39°–40°C [102.2°–104.0°F]) lasting several days, an accompanying generalized eruption of sterile pustules, and a background of intense erythema; patients may become erythrodermic. Episodes of fever and pustules are recurrent. Local irritants, pregnancy, medications, infections, and systemic glucocorticoid withdrawal can precipitate this form of psoriasis. Oral retinoids are the treatment of choice in nonpregnant patients.

Fingernail involvement, appearing as punctate pitting, onycholysis, nail thickening, or subungual hyperkeratosis, may be a clue to the diagnosis of psoriasis when the clinical presentation is not classic.

According to the National Psoriasis Foundation, up to 30% of patients with psoriasis have psoriatic arthritis (PsA). It develops most commonly between the ages of 30 and 50 years. There are five subtypes of PsA: symmetric PsA, asymmetric PsA, distal PsA, spondylitis, and arthritis mutilans. Approximately 50% of PsA is classified as symmetric, which may resemble rheumatoid arthritis. Asymmetric arthritis comprises about 35% of cases. It can involve any joint and may present as “sausage digits.” Distal PsA is the classic form; however, it occurs in only about 5% of patients with PsA. It can involve fingers and toes; fingernails and toenails are often dystrophic, including nail pitting. Spondylitis also occurs in ~5% of patients with PsA. Arthritis mutilans is severe and deforming and affects primarily the small joints of the hands and feet. It accounts for fewer than 5% of PsA cases.

An increased risk of metabolic syndrome, including increased morbidity and mortality from cardiovascular events, has been demonstrated in psoriasis patients. Appropriate screening tests should be performed. The etiology of psoriasis is still poorly understood, but there is clearly a genetic component to the disease. In various studies, 30–50% of patients with psoriasis report a positive family history. Psoriatic lesions contain infiltrates of activated T cells that are thought to elaborate cytokines responsible for keratinocyte hyperproliferation, which results in the characteristic clinical findings. Agents inhibiting T-cell activation, clonal expansion, or release of proinflammatory cytokines are often effective for the treatment of severe psoriasis (see below).

TREATMENT

Psoriasis

Treatment of psoriasis depends on the type, location, and extent of disease. All patients should be instructed to avoid excess drying or irritation of their skin and to maintain adequate cutaneous hydration. Most cases of localized, plaque-type psoriasis can be managed

with mid-potency topical glucocorticoids, although their long-term use is often accompanied by loss of effectiveness (tachyphylaxis) and atrophy of the skin. A topical vitamin D analogue (calcipotriene) and a retinoid (tazarotene) are also efficacious in the treatment of limited psoriasis and have largely replaced other topical agents such as coal tar, salicylic acid, and anthralin.

Ultraviolet (UV) light, natural or artificial, is an effective therapy for many patients with widespread psoriasis. Ultraviolet B (UVB), narrowband UVB, and ultraviolet A (UVA) light with either oral or topical psoralens (PUVA) are used clinically. UV light's immunosuppressive properties are thought to be responsible for its therapeutic activity in psoriasis. It is also mutagenic, potentially leading to an increased incidence of nonmelanoma and melanoma skin cancer. UV-light therapy is contraindicated in patients receiving cyclosporine and should be used with great care in all immunocompromised patients due to the increased risk of skin cancer.

Various systemic agents can be used for severe, widespread psoriatic disease (**Table 57-3**). Oral glucocorticoids should not be used for the treatment of psoriasis due to the potential for development of life-threatening pustular psoriasis when therapy is discontinued. Methotrexate is an effective agent, especially in patients with PsA. The synthetic retinoid acitretin is useful, especially when immunosuppression must be avoided; however, teratogenicity limits its use. Apremilast inhibits phosphodiesterase type 4. It is approved for both psoriasis and PsA. It must be used cautiously in the presence of renal failure or depression.

The evidence implicating psoriasis as a T-cell-mediated disorder has directed therapeutic efforts to immunoregulation. Cyclosporine and other immunosuppressive agents can be very effective in the treatment of psoriasis, and much attention is currently directed toward the development of biologic agents with more selective immunosuppressive properties and better safety profiles (**Table 57-4**). These biologic agents appear to be quite efficacious in treatment of psoriasis and are well tolerated; however, caution with certain patient comorbidities must be exercised. Use of tumor necrosis factor- α (TNF- α) inhibitors may worsen congestive heart failure (CHF), and they should be used with caution in patients at risk for or known to have CHF. Further, none of the immunosuppressive agents used in the treatment of psoriasis should be initiated if the patient has a severe infection (including tuberculosis, HIV, hepatitis B or C); patients on such therapy should be routinely screened for tuberculosis. There have been reports of progressive multifocal leukoencephalopathy and lupus erythematosus in association with treatment with the TNF- α inhibitors. Malignancies, including a risk or history of certain malignancies, may limit the use of these systemic agents. In general, immunosuppressive agents have also been linked to an increase risk of skin cancer, and patients receiving these agents should be monitored for the development of skin cancer.

LICHEN PLANUS

Lichen planus (LP) is a papulosquamous disorder that may affect the skin, scalp, nails, and mucous membranes. The primary cutaneous lesions are pruritic, polygonal, flat-topped, violaceous papules. Close

TABLE 57-3 FDA-Approved Systemic Therapy for Psoriasis

AGENT	MEDICATION CLASS	ADMINISTRATION		ADVERSE EVENTS (SELECTED)
		ROUTE	FREQUENCY	
Methotrexate	Antimetabolite	Oral	Weekly ^a	Hepatotoxicity, pulmonary toxicity, pancytopenia, potential for increased malignancies, ulcerative stomatitis, nausea, diarrhea, teratogenicity
Acitretin	Retinoid	Oral	Daily	Teratogenicity, hepatotoxicity, hyperostosis, hyperlipidemia/pancreatitis, depression, ophthalmologic effects, pseudotumor cerebri
Cyclosporine	Calcineurin inhibitor	Oral	Twice daily	Renal dysfunction, hypertension, hyperkalemia, hyperuricemia, hypomagnesemia, hyperlipidemia, increased risk of malignances
Apremilast	Phosphodiesterase type 4 inhibitor	Oral	Twice daily ^b	Hypersensitivity reaction, depression, nausea, diarrhea, vomiting, dyspepsia, weight loss, headache, fatigue

^aInitial test dose is required. ^bInitial dose escalation is required.

Abbreviation: FDA, Food and Drug Administration.

TABLE 57-4 FDA-Approved Biologics for Psoriasis or Psoriatic Arthritis

MECHANISM OF ACTION	AGENTS (INDICATION; ROUTE)	FREQUENCY	WARNINGS, SELECTED
Anti-TNF- α	Etanercept (Ps, PsA; SC) Adalimumab (Ps, PsA; SC) Certolizumab (Ps, PsA; SC) Infliximab (Ps, PsA; IV) Golimumab (PsA; SC)	Ranges from once or twice weekly ^a to every 8 weeks ^a	Serious infections, hepatotoxicity, CHF, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies
Anti-IL-12 and anti-IL-23	Ustekinumab (Ps, PsA; SC)	Every 12 weeks ^a	Serious infections, neurologic events, potential for increased malignancies
Anti-IL-23	Risankizumab (Ps; SC) Tildrakizumab (Ps; SC) Guselkumab (Ps; SC)	Ranges from every 8–12 weeks ^a	Serious infections, headaches
Anti-IL-17	Secukinumab (Ps, PsA; SC) Ixekizumab (Ps; SC) Brodalumab (Ps; SC)	Ranges from every 2–4 weeks ^a	Serious infections, hypersensitivity reaction, inflammatory bowel disease

^aInitial dose modifications required.

Abbreviations: CHF, congestive heart failure; IL, interleukin; IV, intravenous; Ps, psoriasis; PsA, psoriatic arthritis; SC, subcutaneous; TNF- α , tumor necrosis factor- α .

examination of the surface of these papules often reveals a network of gray lines (*Wickham's striae*). The skin lesions may occur anywhere but have a predilection for the wrists, shins, lower back, and genitalia (Fig. 57-5). Involvement of the scalp (*lichen planopilaris*) may lead to scarring alopecia, and nail involvement may lead to permanent deformity or loss of fingernails and toenails. LP commonly involves mucous membranes, particularly the buccal mucosa, where it can present on a spectrum ranging from a mild, white, reticulate eruption of the mucosa to a severe, erosive stomatitis. Erosive stomatitis may persist for years and may be linked to an increased risk of oral squamous cell carcinoma. Cutaneous eruptions clinically resembling LP have been observed after administration of numerous drugs, including thiazide diuretics, gold, antimarial agents, penicillamine, and phenothiazines, and in patients with skin lesions of chronic graft-versus-host disease. In addition, LP may be associated with hepatitis C infection. The course of LP is variable, but most patients have spontaneous remissions 6 months to 2 years after the onset of disease. Topical glucocorticoids are the mainstay of therapy.

PITYRIASIS ROSEA

Pityriasis rosea (PR) is a papulosquamous eruption of unknown etiology occurring more commonly in the spring and fall. Its first manifestation is the development of a 2- to 6-cm annular lesion (the herald patch). This is followed in a few days to a few weeks by the appearance of many smaller annular or papular lesions with a predilection to occur on the trunk (Fig. 57-6). The lesions are generally oval, with their long

axis parallel to the skinfold lines. Individual lesions may range in color from red to brown and have a trailing scale. PR shares many clinical features with the eruption of secondary syphilis, but palm and sole lesions are extremely rare in PR and common in secondary syphilis. The eruption tends to be moderately pruritic and lasts 3–8 weeks. Treatment is directed at alleviating pruritus and consists of oral antihistamines; mid-potency topical glucocorticoids; and, in some cases, UVB phototherapy.

CUTANEOUS INFECTIONS (TABLE 57-5)

■ IMPETIGO, ECTHYMA, AND FURUNCULOSIS

Impetigo is a common superficial bacterial infection of skin caused most often by *S. aureus* (Chap. 147) and in some cases by group A β -hemolytic streptococci (Chap. 148). The primary lesion is a superficial pustule that ruptures and forms a characteristic yellow-brown honey-colored crust (see Chap. 148, Fig. 148-3). Lesions may occur on normal skin (primary infection) or in areas already affected by another skin disease (secondary infection). Lesions caused by staphylococci may be tense, clear bullae, and this less common form of the disease is called *bullous impetigo*. Blisters are caused by the production of exfoliative toxin by *S. aureus* phage type II. This is the same toxin responsible for staphylococcal scalded-skin syndrome, often resulting in dramatic loss of the superficial epidermis due to blistering. The latter syndrome is much more common in children than in adults; however, it should be considered along with toxic epidermal necrolysis



FIGURE 57-5 Lichen planus. An example of lichen planus showing multiple flat-topped, violaceous papules and plaques. Nail dystrophy, as seen in this patient's thumbnail, may also be a feature. (Courtesy of Robert Swerlick, MD; with permission.)



FIGURE 57-6 Pityriasis rosea. In this patient with pityriasis rosea, multiple round to oval erythematous patches with fine central scale are distributed along the skin tension lines on the trunk.

TABLE 57-5 Common Skin Infections

	CLINICAL FEATURES	ETIOLOGIC AGENT	TREATMENT
Impetigo	Honey-colored crusted papules, plaques, or bullae	Group A <i>Streptococcus</i> and <i>Staphylococcus aureus</i>	Systemic or topical antistaphylococcal and antistreptococcal antibiotics
Dermatophytosis	Inflammatory or noninflammatory annular scaly plaques; may involve hair loss; groin involvement spares scrotum; hyphae on KOH preparation	<i>Trichophyton</i> , <i>Epidermophyton</i> , or <i>Microsporum</i> spp.	Topical azoles, systemic griseofulvin, terbinafine, or azoles
Candidiasis	Inflammatory papules and plaques with satellite pustules, frequently in intertriginous areas; may involve scrotum; pseudohyphae on KOH preparation	<i>Candida albicans</i> and other <i>Candida</i> spp.	Topical nystatin or azoles; systemic azoles for resistant disease
Tinea versicolor	Hyper- or hypopigmented scaly patches on trunk; characteristic mixture of hyphae and spores ("spaghetti and meatballs") on KOH preparation	<i>Malassezia furfur</i>	Topical selenium sulfide lotion or azoles

Abbreviation: KOH, potassium hydroxide.

and severe drug eruptions in patients with widespread blistering of the skin. *Ecthyma* is a deep nonbullous variant of impetigo that causes punched-out ulcerative lesions. It is more often caused by a primary or secondary infection with *Streptococcus pyogenes*. Ecthyma is a deeper infection than typical impetigo and resolves with scars. Treatment of both ecthyma and impetigo involves gentle debridement of adherent crusts, facilitated by using soaks and topical antibiotics in conjunction with appropriate oral antibiotics.

Furunculosis is also caused by *S. aureus*, and this disorder has gained prominence in the past few decades because of CA-MRSA. A furuncle, or boil, is a painful, erythematous nodule that can occur on any cutaneous surface. The lesions may be solitary but are most often multiple. Patients frequently believe they have been bitten by spiders or insects. Family members or close contacts may also be affected. Furuncles can rupture and drain spontaneously or may need incision and drainage, which may be adequate therapy for small solitary furuncles without cellulitis or systemic symptoms. Whenever possible, lesional material should be sent for culture. Current recommendations for methicillin-sensitive infections are β -lactam antibiotics. Therapy for CA-MRSA is discussed previously (see "Atopic Dermatitis"). Warm compresses and nasal mupirocin are helpful therapeutic additions. Severe infections may require IV antibiotics.

■ ERYSIPelas AND CELLULITIS

See Chap. 129.

■ DERMATOPHYTOSIS

Dermatophytes are fungi that infect skin, hair, and nails and include members of the genera *Trichophyton*, *Microsporum*, and *Epidermophyton* (Chap. 219). *Tinea corporis*, or infection of the relatively hairless skin of the body (glabrous skin), may have a variable appearance depending on the extent of the associated inflammatory reaction. Typical infections consist of erythematous, scaly plaques, with an annular appearance that accounts for the common name "ringworm." Deep inflammatory nodules or granulomas occur in some infections, most often those inadequately treated with mid- to high-potency topical glucocorticoids. Involvement of the groin (*tinea cruris*) is more common in males than in females. It presents as a scaling, erythematous eruption sparing the scrotum. Infection of the foot (*tinea pedis*) is the most common dermatophyte infection and is often chronic; it is characterized by variable erythema, edema, scaling, pruritus, and occasionally vesication. The infection may be widespread or localized but generally involves the web space between the fourth and fifth toes. Infection of the nails (*tinea unguium* or *onychomycosis*) occurs in many patients with *tinea pedis* and is characterized by opacified, thickened nails and subungual debris. The distal-lateral variant is most common. Proximal subungual onychomycosis may be a marker for HIV infection or other immunocompromised states. Dermatophyte infection of the scalp (*tinea capitis*) continues to be common, particularly affecting inner-city children but also affecting immunocompromised adults. The predominant organism is *Trichophyton tonsurans*, which can produce a relatively noninflammatory infection with mild scale and hair

loss that is diffuse or localized. *T. tonsurans* and *Microsporum canis* can also cause a markedly inflammatory dermatosis with edema and nodules. This latter presentation is a *kerion*.

The diagnosis of tinea can be made from skin scrapings, nail scrapings, or hair by culture or direct microscopic examination with KOH. Nail clippings may be sent for histologic examination with periodic acid-Schiff (PAS) stain.

TREATMENT

Dermatophytosis

Both topical and systemic therapies may be used in dermatophyte infections. Treatment depends on the site involved and the type of infection. Topical therapy is generally effective for uncomplicated *tinea corporis*, *tinea cruris*, and limited *tinea pedis*. Topical agents are not effective as monotherapy for *tinea capitis* or *onychomycosis* (see below), and nystatin is not active against dermatophytes. Topicals are generally applied twice daily, and treatment should continue for 1 week beyond clinical resolution of the infection. *Tinea pedis* often requires longer treatment courses and frequently relapses. Oral antifungal agents may be required for recalcitrant *tinea pedis* or *tinea corporis*.

For dermatophyte infections involving the hair and nails and for other infections unresponsive to topical therapy, oral antifungal agents are often used. Markedly inflammatory *tinea capitis* may result in scarring and hair loss, and a systemic antifungal agent plus systemic or topical glucocorticoids may be helpful in preventing these sequelae. A fungal etiology should be confirmed by direct microscopic examination or by culture before oral antifungal agents are prescribed for any infection. All the oral agents may cause hepatotoxicity. They should not be used in women who are pregnant or breast-feeding.

Griseofulvin is approved in the United States for dermatophyte infections involving the skin, hair, or nails. Common side effects of griseofulvin include gastrointestinal distress, headache, and urticaria.

Two other oral antifungal agents, itraconazole and terbinafine, are sometimes prescribed "off-label" for superficial fungal infections. Oral itraconazole is approved for onychomycosis. Itraconazole has the potential for serious interactions with other drugs requiring the P450 enzyme system for metabolism. Itraconazole should not be administered to patients with evidence of ventricular dysfunction or patients with known CHF.

Terbinafine is also approved for onychomycosis, and the granule version is approved for treatment of *tinea capitis*. Terbinafine has fewer interactions with other drugs than itraconazole; however, caution should be used with patients who are on multiple medications. The risk/benefit ratio should be considered when an asymptomatic toenail infection is treated with systemic agents.

The FDA has limited the use of a third oral agent due to potential hepatotoxicity and published the following: "Nizoral [ketoconazole]

oral tablets should not be a first-line treatment for any fungal infection." The topical form of ketoconazole is not affected by this action.

TINEA (PITYRIASIS) VERSICOLOR

Tinea versicolor is caused by a nondermatophytic, dimorphic fungus, *Malassezia furfur*, a normal inhabitant of the skin. The expression of infection is promoted by heat and humidity. The typical lesions consist of oval scaly macules, papules, and patches concentrated on the chest, shoulders, and back but only rarely on the face or distal extremities. On dark skin, the lesions often appear as hypopigmented areas, whereas on light skin, they are slightly erythematous or hyperpigmented. A KOH preparation from scaling lesions will demonstrate a confluence of short hyphae and round spores ("spaghetti and meatballs"). Lotions or shampoos containing sulfur, salicylic acid, or selenium sulfide are the treatments of choice and will clear the infection if used daily for 1–2 weeks and then weekly thereafter. These preparations are irritating if left on the skin for >10 min; thus, they should be washed off completely. Treatment with some oral antifungal agents is also effective, but they do not provide lasting results and are not FDA approved for this indication.

CANDIDIASIS

Candidiasis is a fungal infection caused by a related group of yeasts whose manifestations may be localized to the skin and mucous membranes or, rarely, may be systemic and life-threatening (Chap. 216). The causative organism is usually *Candida albicans*. These organisms are normal saprophytic inhabitants of the gastrointestinal tract but may overgrow due to broad-spectrum antibiotic therapy, diabetes mellitus, or immunosuppression and cause disease. Candidiasis is a very common infection in HIV-infected individuals (Chap. 202). The oral cavity is commonly involved. Lesions may occur on the tongue or buccal mucosa (*thrush*) and appear as white plaques. Fissured, macerated lesions at the corners of the mouth (*perlèche*) are often seen in individuals with poorly fitting dentures and may also be associated with candidal infection. In addition, candidal infections have an affinity for sites that are chronically wet and macerated, including the skin around nails (onycholysis and paronychia), and in intertriginous areas. Intertriginous lesions are characteristically edematous, erythematous, and scaly, with scattered "satellite pustules." In males, there is often involvement of the penis and scrotum as well as the inner aspect of the thighs. In contrast to dermatophyte infections, candidal infections are frequently painful and accompanied by a marked inflammatory response. Diagnosis of candidal infection is based on the clinical pattern and demonstration of yeast on KOH preparation or culture.

TREATMENT

Candidiasis

Treatment involves removal of any predisposing factors such as antibiotic therapy or chronic moisture and the use of appropriate topical or systemic antifungal agents. Effective topicals include nystatin or azoles (miconazole, clotrimazole, econazole, or ketoconazole). The associated inflammatory response accompanying candidal infection on glabrous skin can be treated with a mild glucocorticoid lotion or cream (2.5% hydrocortisone). Systemic therapy is usually reserved for immunosuppressed patients or individuals with chronic or recurrent disease who fail to respond to appropriate topical therapy. Oral fluconazole is most commonly prescribed for cutaneous candidiasis. Oral nystatin is effective only for candidiasis of the gastrointestinal tract.

WARTS

Warts are cutaneous neoplasms caused by papillomaviruses. More than 100 different human papillomaviruses (HPVs) have been described. A typical wart, *verruca vulgaris*, is sessile, dome-shaped, and usually about a centimeter in diameter. Its surface is hyperkeratotic, consisting of many small filamentous projections. HPV also causes typical planter warts, flat warts (*verruca plana*), and filiform warts. Plantar warts

are endophytic and are covered by thick keratin. Paring of the wart will generally reveal a central core of keratinized debris and punctate bleeding points. Filiform warts are commonly seen on the face, neck, and skinfolds and present as papillomatous lesions on a narrow base. Flat warts are only slightly elevated and have a velvety, non verrucous surface. They have a propensity for the face, arms, and legs, and are often spread by shaving.

Genital warts begin as small papillomas that may grow to form large, fungating lesions. In women, they may involve the labia, perineum, or perianal skin. In addition, the mucosa of the vagina, urethra, and anus can be involved as well as the cervical epithelium. In men, the lesions often occur initially in the coronal sulcus but may be seen on the shaft of the penis, the scrotum, or the perianal skin or in the urethra.

Appreciable evidence has accumulated indicating that HPV plays a role in the development of neoplasia of the uterine cervix and anogenital skin (Chap. 89). HPV types 16 and 18 have been most intensely studied and are the major risk factors for intraepithelial neoplasia and squamous cell carcinoma of the cervix, anus, vulva, and penis. The risk is higher among patients immunosuppressed after solid organ transplantation and among those infected with HIV. Recent evidence also implicates other HPV types. Histologic examination of biopsied samples from affected sites may reveal changes associated with typical warts and/or features typical of intraepidermal carcinoma (Bowen's disease). Squamous cell carcinomas associated with HPV infections have also been observed in extragenital skin (Chap. 76), most commonly in patients immunosuppressed after organ transplantation. Patients on long-term immunosuppression should be monitored for the development of squamous cell carcinoma and other cutaneous malignancies.

TREATMENT

Warts

Treatment of warts, other than anogenital warts, should be tempered by the observation that most warts in normal individuals resolve spontaneously within 1–2 years. There are many modalities available to treat warts, but no single therapy is universally effective. Factors that influence the choice of therapy include the location of the wart, the extent of disease, the age and immunologic status of the patient, and the patient's desire for therapy. Perhaps the most useful and convenient method for treating warts in almost any location is cryotherapy with liquid nitrogen. Equally effective for non-genital warts, but requiring much more patient compliance, is the use of keratolytic agents such as salicylic acid plasters or solutions. For genital warts, in-office application of a podophyllin solution is moderately effective but may be associated with marked local reactions. Prescription preparations of dilute, purified podophyllin are available for home use. Topical imiquimod, a potent inducer of local cytokine release, has been approved for treatment of genital warts. A topical compound composed of green tea extracts (sinecatechins) is also available. Conventional and laser surgical procedures may be required for recalcitrant warts. Recurrence of warts appears to be common with all these modalities. A highly effective vaccine for selected types of HPV has been approved by the FDA, and its use is reported to reduce the incidence of anogenital and cervical carcinoma.

HERPES SIMPLEX

See Chap. 192.

HERPES ZOSTER

See Chap. 193.

ACNE

ACNE VULGARIS

Acne vulgaris is a self-limited disorder primarily of teenagers and young adults, although perhaps 10–20% of adults may continue to experience some form of the disorder. The permissive factor for the

expression of the disease in adolescence is the increase in sebum production by sebaceous glands with puberty. Small cysts, called *comedones*, form in hair follicles due to blockage of the follicular orifice by retention of keratinous material and sebum. The activity of bacteria (*Cutibacterium acnes*) within the comedones releases free fatty acids from sebum, causes inflammation within the cyst, and results in rupture of the cyst wall. An inflammatory foreign-body reaction develops as result of extrusion of oily and keratinous debris from the cyst.

The clinical hallmark of acne vulgaris is the comedone, which may be closed (*whitehead*) or open (*blackhead*). Closed comedones appear as 1- to 2-mm pebbly white papules, which are accentuated when the skin is stretched. They are the precursors of inflammatory lesions of acne vulgaris. The contents of closed comedones are not easily expressed. Open comedones, which rarely result in inflammatory acne lesions, have a dilated follicular orifice and are filled with easily expressible oxidized, darkened, oily debris. Comedones are usually accompanied by inflammatory lesions: papules, pustules, or nodules.

The earliest lesions seen in adolescence are generally mildly inflamed or noninflammatory comedones on the forehead. Subsequently, more typical inflammatory lesions develop on the cheeks, nose, and chin (Fig. 57-7). The most common location for acne is the face, but involvement of the chest and back is common. Most disease remains mild and does not lead to scarring. A small number of patients develop large inflammatory cysts and nodules, which may drain and result in significant scarring. Regardless of the severity, acne may affect a patient's quality of life. With adequate treatment, this effect may be transient. In the case of severe, scarring acne, the effects can be permanent and profound. Early therapeutic intervention in severe acne is essential.

Exogenous and endogenous factors can alter the expression of acne vulgaris. Friction and trauma (from headbands or chin straps of athletic helmets), application of comedogenic topical agents (cosmetics or hair preparations), or chronic topical exposure to certain industrial compounds may elicit or aggravate acne. Glucocorticoids, topical or systemic, may also elicit acne. Other systemic medications such as progestin-only contraception, lithium, isoniazid, androgenic steroids, halogens, phenytoin, and phenobarbital may produce acneiform eruptions or aggravate preexisting acne. Genetic factors and polycystic ovary disease may also play a role.

TREATMENT

Acne Vulgaris

Treatment of acne vulgaris is directed toward elimination of comedones by normalizing follicular keratinization and decreasing sebaceous gland activity, the population of *C. acnes*, and inflammation. Minimal to moderate pauci-inflammatory disease may respond adequately to local therapy alone. Although areas affected with acne should be kept clean, overly vigorous scrubbing may aggravate acne due to mechanical rupture of comedones. Topical agents such as retinoic acid, benzoyl peroxide, or salicylic acid may alter



FIGURE 57-7 Acne vulgaris. An example of acne vulgaris with inflammatory papules, pustules, and comedones. (Courtesy of Kalman Watsky, MD; with permission.)

the pattern of epidermal desquamation, preventing the formation of comedones and aiding in the resolution of preexisting cysts. Topical antibacterial agents (such as benzoyl peroxide, azelaic acid, erythromycin, clindamycin, or dapsone) are also useful adjuncts to therapy. Topical antibiotics (erythromycin and clindamycin) should be used in combination with benzoyl peroxide to prevent development of bacterial resistance.

Patients with moderate to severe acne with a prominent inflammatory component will benefit from the addition of systemic therapy, such as minocycline or doxycycline in doses of 100 mg bid or in lower dose, extended-release preparations. Such antibiotics appear to have anti-inflammatory effects independent of their antibacterial effects. Female patients who do not respond to oral antibiotics may benefit from hormonal therapy. Several oral contraceptives are now approved by the FDA for use in the treatment of acne vulgaris. Spironolactone is emerging as a safe, effective, and durable antian-rogen treatment in women.

Patients with severe nodulocystic acne unresponsive to the therapies discussed above may benefit from treatment with the synthetic retinoid isotretinoin. Dosing is weight-based and cumulative, with duration of therapy dictated by summative dose or acne lesion remission. Results are excellent in appropriately selected patients. Its use is highly regulated due to its potential for severe adverse events, primarily teratogenicity and depression. In addition, patients receiving this medication develop dry skin and cheilitis and must be followed for development of hypertriglyceridemia.

At present, prescribers must enroll in a program designed to prevent pregnancy and adverse events while patients are taking isotretinoin. These measures are imposed to ensure that all prescribers are familiar with the risks of isotretinoin, that all female patients have two negative pregnancy tests prior to initiation of therapy and a negative pregnancy test prior to each refill, and that all patients have been warned about the risks associated with isotretinoin.

ACNE ROSACEA

Acne rosacea, commonly referred to simply as *rosacea*, is an inflammatory disorder predominantly affecting the central face. Persons most often affected are Caucasians of northern European background, but rosacea also occurs in patients with dark skin. Rosacea is seen almost exclusively in adults, only rarely affecting patients <30 years old. Rosacea is more common in women, but those most severely affected are men. It is characterized by the presence of erythema, telangiectasias, and superficial pustules (Fig. 57-8) but is not associated with the presence of comedones. Rosacea rarely involves the chest or back.

There is a relationship between the tendency for facial flushing and the subsequent development of acne rosacea. Often, individuals with rosacea initially demonstrate a pronounced flushing reaction. This may be in response to heat, emotional stimuli, alcohol, hot drinks, or spicy foods. As the disease progresses, the flush persists longer and longer and may eventually become permanent. Papules, pustules, and



FIGURE 57-8 Acne rosacea. Prominent facial erythema, telangiectasias, scattered papules, and small pustules are seen in this patient with acne rosacea. (Courtesy of Robert Swerlick, MD; with permission.)

telangiectasias can become superimposed on the persistent flush. Rosacea of very long standing may lead to connective tissue overgrowth, particularly of the nose (*rhinophyma*). Rosacea may also be complicated by various inflammatory disorders of the eye, including keratitis, blepharitis, iritis, and recurrent chalazion. These ocular problems are potentially sight-threatening and warrant ophthalmologic evaluation.

TREATMENT

Acne Rosacea

Acne rosacea can be treated topically or systemically. Mild disease often responds to topical preparations of metronidazole, sodium sulfacetamide, azelaic acid, ivermectin, brimonidine, or oxymetazoline. More severe disease requires oral tetracyclines in subantimicrobial, modified-release preparations. Residual telangiectasia may respond to laser therapy. Topical glucocorticoids, especially potent agents, should be avoided because chronic use of these preparations may elicit rosacea. Application of topical agents to the skin is not effective treatment for ocular disease.

SKIN DISEASES AND SMALLPOX VACCINATION

Although smallpox vaccinations were discontinued several decades ago for the general population, they are still required for certain military personnel and first responders. In the absence of a bioterrorism attack and a real or potential exposure to smallpox, such vaccination is contraindicated in persons with a history of skin diseases, such as AD, eczema, and psoriasis, who have a higher incidence of adverse events associated with smallpox vaccination. In the case of such exposure, the risk of smallpox infection outweighs that of adverse events from the vaccine (Chap. S3).

FURTHER READING

- BOLOGNA JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.
 JAMES WD et al (eds): *Andrew's Diseases of the Skin Clinical Dermatology*, 13th ed. Philadelphia, Elsevier, 2020.
 KANG S et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 9th ed. New York, McGraw-Hill, 2019.
 WOLFF K et al (eds): *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 8th ed. New York, McGraw-Hill, 2017.

individual diseases, but by describing the various presenting clinical signs and symptoms that point to specific disorders. Concise differential diagnoses will be generated in which the significant diseases will be distinguished from the more common cutaneous disorders that have minimal or no significance with regard to associated internal disease. The latter disorders are reviewed in table form and always need to be excluded when considering the former. For a detailed description of individual diseases, the reader should consult a dermatologic text.

PAPULOSQUAMOUS SKIN LESIONS

(Table 58-1) When an eruption is characterized by elevated lesions, either papules (<1 cm) or plaques (>1 cm), in association with scale, it is referred to as *papulosquamous*. The most common papulosquamous diseases—*tinea*, *psoriasis*, *pityriasis rosea*, and *lichen planus*—are primary cutaneous disorders (Chap. 57). When psoriatic lesions are accompanied by arthritis, the possibility of psoriatic arthritis or reactive arthritis should be considered. A history of oral ulcers, conjunctivitis, uveitis, and/or urethritis points to the latter diagnosis. Lithium, beta blockers, anti-PD-1/PD-L1 antibodies, HIV or streptococcal infections, and a rapid taper of systemic glucocorticoids are known to exacerbate psoriasis; despite being used to treat psoriasis, tumor necrosis factor (TNF) inhibitors can also induce psoriatic lesions. Comorbidities in patients with psoriasis include cardiovascular disease and metabolic syndrome.

Whenever the clinical diagnosis of pityriasis rosea or lichen planus is made, it is important to review the patient's medications because the eruption may resolve by simply discontinuing the offending agent. Pityriasis rosea-like drug eruptions are seen most commonly with beta blockers, angiotensin-converting enzyme (ACE) inhibitors, and metronidazole, whereas the drugs that can produce a lichenoid eruption include thiazides, antimalarials, quinidine, beta blockers, TNF inhibitors, anti-PD-1/PD-L1 antibodies, and ACE inhibitors. In some populations (e.g., Europeans), there is a higher prevalence of hepatitis C viral infection in patients with oral lichen planus. Lichen planus-like lesions are also observed in chronic graft-versus-host disease.

In its early stages, the mycosis fungoïdes (MF) form of *cutaneous T-cell lymphoma* (CTCL) may be confused with eczema or psoriasis, but it often eventually fails to respond to appropriate therapy for those inflammatory diseases. MF can develop within lesions of large-plaque parapsoriasis and is suggested by an increase in the thickness of the lesions. The diagnosis of MF is established by skin biopsy in which

TABLE 58-1 Selected Causes of Papulosquamous Skin Lesions

1. Primary cutaneous disorders
 - a. *Tinea*^a—widespread disease may be sign of immunosuppression
 - b. *Psoriasis*^a—widespread or resistant disease may be sign of HIV infection
 - c. *Pityriasis rosea*^a
 - d. *Lichen planus*^a
 - e. *Parapsoriasis*, small plaque and large plaque
 - f. *Bowen's disease* (squamous cell carcinoma *in situ*)^b
2. Drugs
3. Systemic diseases
 - a. *Lupus erythematosus*, primarily subacute or chronic (discoid) lesions^c
 - b. *Cutaneous T-cell lymphoma*, in particular, *mycosis fungoïdes*^d
 - c. *Secondary syphilis*
 - d. *Reactive arthritis*
 - e. *Sarcoidosis*^e—with scale less common than without scale
 - f. *Bazex syndrome* (*acrokeratosis paraneoplastica*)^f

^aDiscussed in detail in Chap. 57; cardiovascular disease and the metabolic syndrome are comorbidities in psoriasis; primarily in Europe, hepatitis C virus is associated with oral lichen planus. ^bAssociated with chronic sun exposure more often than exposure to arsenic; usually one or a few lesions. ^cSee also Red Lesions in "Papulonodular Skin Lesions." ^dAlso cutaneous lesions of HTLV-1-associated adult T-cell leukemia/lymphoma. ^eSee also Red-Brown Lesions in "Papulonodular Skin Lesions." ^fPsoriasisiform lesions of the helices, nose, and acral sites; squamous cell carcinoma of the upper aerodigestive tract most common underlying malignancy.

Abbreviation: HIV, human immunodeficiency virus.

58

Skin Manifestations of Internal Disease

Jean L. Bolognia, Jonathan S. Leventhal,
 Irwin M. Braverman

It is a generally accepted concept in medicine that the skin can develop signs of internal disease. Therefore, in textbooks of medicine, one finds a chapter describing in detail the major systemic disorders that can be identified by cutaneous signs. The underlying assumption of such a chapter is that the clinician has been able to identify the specific disorder in the patient and needs only to read about it in the textbook. In reality, concise differential diagnoses and the identification of these disorders are actually difficult for the nondermatologist because he or she is not well-versed in the recognition of cutaneous lesions or their spectrum of presentations. Therefore, this chapter covers this particular topic of cutaneous medicine not by simply focusing on



TABLE 58-2 Causes of Erythroderma

1. Primary cutaneous disorders
 - a. Psoriasis^a
 - b. Dermatitis (atopic > contact > stasis [with autosensitization] or seborrheic [primarily infants])^a
 - c. Pityriasis rubra pilaris
2. Drugs
3. Systemic diseases
 - a. Cutaneous T-cell lymphoma (Sézary syndrome, erythrodermic mycosis fungoïdes)
 - b. Other lymphomas
 - c. Rarely, late-stage solid tumors
4. Idiopathic (usually older men)

^aDiscussed in detail in **Chap. 57**.

collections of atypical T lymphocytes are found in the epidermis and dermis. As the disease progresses, cutaneous tumors and lymph node involvement may appear.

In *secondary syphilis*, there are scattered pink to red-brown papules with thin scale. The eruption often involves the palms and soles and can resemble pityriasis rosea. Associated findings are helpful in making the diagnosis and include nonscarring alopecia, annular plaques on the face, mucous patches, condyloma lata (broad-based and moist), and lymphadenopathy, as well as malaise, fever, headache, and myalgias. The interval between the primary chancre and the secondary stage is usually 4–8 weeks, and spontaneous resolution without appropriate therapy occurs.

ERYthroderMA

(Table 58-2) *Erythroderma* is the term used when the majority of the skin surface is erythematous (red in color). There may be associated scale, erosions, or pustules as well as shedding of the hair and nails. Potential systemic manifestations include fever, chills, hypothermia, reactive lymphadenopathy, peripheral edema, hypoalbuminemia, and high-output cardiac failure. The major etiologies of erythroderma are (1) cutaneous diseases such as psoriasis and dermatitis (Table 58-3); (2) drugs; (3) systemic diseases, most commonly CTCL; and (4) idiopathic. In the first three groups, the location and description of the initial lesions, prior to the development of the erythroderma, aid in the diagnosis. For example, a history of red scaly plaques on the elbows and knees would point to psoriasis. It is also important to examine the skin carefully for a migration of the erythema and associated secondary changes such as pustules or erosions. Migratory waves of erythema studded with superficial pustules are seen in *pustular psoriasis*.

Drug-induced erythroderma may begin as an exanthematos (morbilliform) eruption (Chap. 60) or may arise as diffuse erythema. A number of drugs can produce an erythroderma, including penicillins, sulfonamides, aromatic anticonvulsants (e.g., carbamazepine, phenytoin), and allopurinol. Fever and peripheral eosinophilia often accompany the eruption, and there may also be facial swelling, hepatitis, myocarditis, thyroiditis, and allergic interstitial nephritis; this constellation is frequently referred to as *drug reaction with eosinophilia and systemic symptoms* (DRESS) or *drug-induced hypersensitivity syndrome* (DIHS). In addition, these reactions, especially to aromatic anticonvulsants, can lead to a pseudolymphoma syndrome with adenopathy and circulating atypical lymphocytes, while reactions to allopurinol may be accompanied by gastrointestinal bleeding.

The most common malignancy that is associated with erythroderma is CTCL; in some series, up to 25% of the cases of erythroderma were due to CTCL. The patient may progress from isolated plaques and tumors, but more commonly, the erythroderma is present throughout the course of the disease (Sézary syndrome). In Sézary syndrome, there are circulating clonal atypical T lymphocytes, pruritus, and lymphadenopathy. In cases of erythroderma where there is no apparent cause (idiopathic), longitudinal evaluation is mandatory to monitor for the possible development of CTCL.

ALOPECIA

(Table 58-4) The two major forms of alopecia are scarring and non-scarring. *Scarring alopecia* is associated with fibrosis, inflammation, and loss of hair follicles. A smooth scalp with a decreased number of follicular openings is usually observed clinically, but in some patients, the changes are seen only in biopsy specimens from affected areas. In *nonscarring alopecia*, the hair shafts are absent or miniaturized, but the hair follicles are preserved, explaining the reversible nature of nonscarring alopecia.

The most common causes of nonscarring alopecia include *androgenetic alopecia*, *telogen effluvium*, *alopecia areata*, *tinea capitis*, and the early phase of *traumatic alopecia* (Table 58-5). In women with androgenetic alopecia, an elevation in circulating levels of androgens may be seen as a result of ovarian or adrenal gland dysfunction or neoplasm. When there are signs of virilization, such as a deepened voice and/or enlarged clitoris, the possibility of an ovarian or adrenal gland tumor should be considered.

Exposure to various drugs can also cause diffuse hair loss, usually by inducing a telogen effluvium. An exception is the anagen effluvium observed with chemotherapeutic agents such as daunorubicin. Alopecia is a side effect of the following drugs: warfarin, heparin, propylthiouracil, carbimazole, isotretinoin, acitretin, lithium, beta blockers, interferons, colchicine, and amphetamines. Fortunately, spontaneous regrowth usually follows discontinuation of the offending agent.

Less commonly, nonscarring alopecia is associated with *lupus erythematosus* and *secondary syphilis*. In systemic lupus, there are two forms of alopecia—one is scarring secondary to discoid lesions (see below), and the other is nonscarring. The latter form coincides with flares of systemic disease and may involve the entire scalp or just the frontal scalp, with the appearance of multiple short hairs ("lupus hairs") as a sign of initial regrowth. Scattered, poorly circumscribed patches of alopecia with a "moth-eaten" appearance are a manifestation of the secondary stage of syphilis. Diffuse thinning of the hair is also associated with hypothyroidism and hyperthyroidism (Table 58-4).

Scarring alopecia is more frequently the result of a primary cutaneous disorder such as *lichen planus*, *chronic cutaneous (discoid) lupus*, *central centrifugal cicatricial alopecia*, *folliculitis decalvans*, or *linear scleroderma (morphia)* than it is a sign of systemic disease. Although the scarring lesions of *discoid lupus* can be seen in patients with systemic lupus, in the majority of patients, the disease process is limited to the skin. Less common causes of scarring alopecia include *sarcoidosis* (see "Papulonodular Skin Lesions," below), chemotherapeutic agents, and *cutaneous metastases*.

In the early phases of discoid lupus, lichen planus, and folliculitis decalvans, there are circumscribed areas of alopecia. Fibrosis and subsequent loss of hair follicles are observed primarily in the center of these alopecic patches, whereas the inflammatory process is most prominent at the periphery. The areas of active inflammation in discoid lupus are erythematous with scale, whereas the areas of previous inflammation are often hypopigmented with a rim of hyperpigmentation. In lichen planus, perifollicular macules at the periphery are usually violet-colored. A complete examination of the skin and oral mucosa combined with a biopsy and direct immunofluorescence microscopy of inflamed skin will aid in distinguishing these two entities. The peripheral active lesions in folliculitis decalvans are follicular pustules; these patients can develop a reactive arthritis.

FIGURATE SKIN LESIONS

(Table 58-6) In *figurate eruptions*, the lesions form rings and arcs that are usually erythematous but can be skin-colored to brown. Most commonly, they are due to primary cutaneous diseases such as *tinea*, *urticaria*, *granuloma annulare*, and *erythema annulare centrifugum* (Chaps. 57 and 59). An underlying systemic illness is found in a second, less common group of migratory annular erythemas. It includes *erythema migrans*, *erythema gyratum repens*, *erythema marginatum*, and *necrolytic migratory erythema*.

In *erythema gyratum repens*, one sees numerous mobile concentric arcs and wavefronts that resemble the grain in wood. A search for an

TABLE 58-3 Erythroderma (Primary Cutaneous Disorders)

	INITIAL LESIONS	LOCATION OF INITIAL LESIONS	OTHER FINDINGS	DIAGNOSTIC AIDS	TREATMENT
Psoriasis ^a	Pink-red, silvery scale, sharply demarcated	Elbows, knees, scalp, presacral area, intergluteal fold	Nail dystrophy (e.g., pits, oil drop sign), arthritis, pustules, SAPHO syndrome ^b	Skin biopsy	Topical glucocorticoids, vitamin D analogs; UV-B (narrowband) > PUVA; oral retinoids; MTX; anti-TNF agents, anti-IL-12/23 Ab, anti-IL-23 Ab, anti-IL-17A or -IL-17 receptor A Ab; apremilast; cyclosporine
Dermatitis^a					
Atopic	Acute: Erythema, fine scale, crust, indistinct borders, excoriations Chronic: Lichenification (increased skin markings), excoriations	Antecubital and popliteal fossae, neck, hands, eyelids	Pruritus Personal and/or family history of atopy, including asthma, allergic rhinitis or conjunctivitis, and atopic dermatitis Exclude secondary infection with <i>Staphylococcus aureus</i> or HSV Exclude superimposed irritant or allergic contact dermatitis	Skin biopsy	Topical glucocorticoids, tacrolimus, pimecrolimus, tar, crisaborole, and antipruritics; oral antihistamines for sedation; open wet dressings; UV-B ± UV-A > PUVA; anti-IL-4/13 Ab; oral/IM glucocorticoids (short-term); MTX; mycophenolate mofetil; azathioprine; cyclosporine Topical or oral antibiotics
Contact	Local: Erythema, crusting, vesicles, and bullae Systemic: Erythema, fine scale, crust	Depends on offending agent Generalized vs major intertriginous zones (especially groin)	Irritant—onset often within hours Allergic—delayed-type hypersensitivity; lag time of 48 h with rechallenge Patient has history of allergic contact dermatitis to topical agent and then receives systemic medication that is structurally related, e.g., formaldehyde (skin), aspartame (oral)	Patch testing; repeat open application test Patch testing	Remove irritant or allergen; topical glucocorticoids; oral antihistamines; oral/IM glucocorticoids (short-term) Same as local
Seborrheic (rare in adults)	Pink-red to pink-orange, greasy scale	Scalp, nasolabial folds, eyebrows, intertriginous zones	Flares with stress, HIV infection Associated with Parkinson's disease	Skin biopsy	Topical glucocorticoids and imidazoles
Stasis (with autosensitization)	Erythema, crusting, excoriations	Lower extremities	Pruritus, lower extremity edema, varicosities, hemosiderin deposits, lipodermatosclerosis History of venous ulcers, thrombophlebitis, and/or cellulitis Exclude cellulitis Exclude superimposed contact dermatitis, e.g., topical neomycin	Skin biopsy	Topical glucocorticoids; open wet dressings; leg elevation; pressure stockings; pressure wraps if associated ulcers
Pityriasis rubra pilaris	Orange-red (salmon-colored), perifollicular papules	Generalized, but characteristic “skip” areas of normal skin	Wax-like palmoplantar keratoderma Exclude cutaneous T-cell lymphoma	Skin biopsy	Isotretinoin or acitretin; MTX; anti-IL-12/23 Ab, anti-IL-23 Ab, anti-TNF agents, anti-IL-17A or -IL-17 receptor A Ab

^aDiscussed in detail in **Chap. 57**. ^bSAPHO syndrome occurs more commonly in patients with palmoplantar pustulosis than in those with erythrodermic psoriasis.

Abbreviations: Ab, antibody; HSV, herpes simplex virus; IL, interleukin; IM, intramuscular; MTX, methotrexate; PUVA, psoralens plus ultraviolet A irradiation; SAPHO, synovitis, acne, pustulosis, hyperostosis, and osteitis (a subtype is chronic recurrent multifocal osteomyelitis); TNF, tumor necrosis factor; UV-A, ultraviolet A irradiation; UV-B, ultraviolet B irradiation.

underlying malignancy is mandatory in a patient with this eruption. Erythema migrans is the cutaneous manifestation of Lyme disease, which is caused by the spirochete *Borrelia burgdorferi*. In the initial stage (3–30 days after tick bite), a single annular lesion is usually seen, which can expand to ≥10 cm in diameter. Within several days, up to half of the patients develop multiple smaller erythematous lesions at sites distant from the bite. Associated symptoms include fever, headache, photophobia, myalgias, arthralgias, and malar rash. Erythema marginatum is seen in patients with rheumatic fever, primarily on the trunk. Lesions are pink-red in color, flat to minimally elevated, and transient.

There are additional cutaneous diseases that present as annular eruptions but lack an obvious migratory component. Examples include *CTCL*, *subacute cutaneous lupus*, *secondary syphilis*, and *sarcoidosis* (see “Papulonodular Skin Lesions,” below).

ACNE

(**Table 58-7**) In addition to *acne vulgaris* and *acne rosacea*, the two major forms of acne (**Chap. 57**), there are drugs and systemic diseases that can lead to acneiform eruptions.

Patients with the *carcinoid syndrome* have episodes of flushing of the head, neck, and sometimes the trunk. Resultant skin changes of the

TABLE 58-4 Causes of Alopecia

I. Nonscarring alopecia
A. Primary cutaneous disorders
1. Androgenetic alopecia (female pattern, male pattern)
2. Telogen effluvium
3. Alopecia areata
4. Tinea capitis
5. Traumatic alopecia ^a
6. Psoriasiform alopecia, including TNF inhibitor–induced
B. Drugs
1. Telogen effluvium—see text for most common causes
2. Anagen effluvium—chemotherapeutic agents (e.g., anthracyclines)
C. Systemic diseases
1. Systemic lupus erythematosus
2. Secondary syphilis
3. Hypothyroidism
4. Hyperthyroidism
5. Hypopituitarism
6. Deficiencies of protein, biotin, zinc, and perhaps iron
II. Scarring alopecia
A. Primary cutaneous disorders
1. Cutaneous lupus (chronic discoid lesions) ^b
2. Lichen planus, including frontal fibrosing alopecia
3. Central centrifugal cicatricial alopecia
4. Folliculitis decalvans
5. Dissecting cellulitis
6. Linear morphea (linear scleroderma) ^c
B. Drugs
1. Chemotherapeutic agents (e.g., taxanes, busulfan)
C. Systemic diseases
1. Discoid lesions in the setting of systemic lupus erythematosus ^b
2. Sarcoidosis
3. Cutaneous metastases

^aMost patients with trichotillomania or early stages of traction alopecia and some patients with pressure-induced alopecia. ^bWhile the majority of patients with discoid lesions have only cutaneous disease, these lesions do represent one of the criteria in the European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) [2019] and ACR [1982] classification schemes for systemic lupus erythematosus. ^cCan involve underlying muscles and osseous structures, and rarely in linear morphea of the frontal scalp (*en coup de sabre*), there is involvement of the meninges and brain.

face, in particular telangiectasias, may mimic the clinical appearance of erythematotelangiectatic acne rosacea.

PUSTULAR LESIONS

Acneiform eruptions (see “Acne,” above) and *folliculitis* represent the most common pustular dermatoses. An important consideration in the evaluation of follicular pustules is a determination of the associated pathogen, for example, normal flora (culture-negative), *Staphylococcus aureus*, *Pseudomonas aeruginosa* (“hot tub” folliculitis), *Malassezia*, dermatophytes (Majocchi’s granuloma), and *Demodex* spp. Noninfectious forms of folliculitis include HIV- or immunosuppression-associated eosinophilic folliculitis and folliculitis secondary to drugs such as glucocorticoids, lithium, and epidermal growth factor receptor (EGFR) or MEK inhibitors. Administration of high-dose systemic glucocorticoids can result in a widespread eruption of follicular pustules on the trunk, characterized by lesions in the same stage of development. With regard to underlying systemic diseases, nonfollicular-based pustules are a characteristic component of pustular psoriasis (sterile) and can be seen in septic emboli of bacterial or fungal origin (see “Purpura,” below). In patients with acute generalized exanthematous pustulosis (AGEP) due primarily to medications (e.g., cephalosporins), there are large areas of erythema studded with multiple sterile pustules in addition to neutrophilia.

TELANGIECTASIAS

(**Table 58-8**) To distinguish the various types of telangiectasias, it is important to examine the shape and configuration of the dilated blood vessels. *Linear telangiectasias* are seen on the face of patients with *actinically damaged skin* and *acne rosacea*, and they are found on the legs of patients with *venous hypertension* and first appear on the legs in *generalized essential telangiectasia*. Patients with an unusual form of *mastocytosis* (telangiectasia macularis eruptiva perstans) and the *carcinoid syndrome* (see “Acne,” above) also have linear telangiectasias. Lastly, linear telangiectasias are found in areas of cutaneous inflammation. For example, longstanding lesions of discoid lupus frequently have telangiectasias within them.

Poikiloderma is a term used to describe a patch of skin with: (1) reticulated hypo- and hyperpigmentation, (2) wrinkling secondary to epidermal atrophy, and (3) telangiectasias. Poikiloderma does not imply a single disease entity—although it is becoming less common, it is seen in skin damaged by *ionizing radiation* as well as in patients with autoimmune connective tissue diseases, primarily *dermatomyositis* (DM), and rare genodermatoses (e.g., Kindler syndrome).

In *systemic sclerosis* (scleroderma), the dilated blood vessels have a unique configuration and are known as *mat telangiectasias*. The lesions are broad macules that usually measure 2–7 mm in diameter but occasionally are larger. Mats have a polygonal or oval shape, and their erythematous color may appear uniform, but, upon closer inspection, the erythema is the result of delicate telangiectasias. The most common locations for mat telangiectasias are the face, oral mucosa, and hands—peripheral sites that are prone to intermittent ischemia. The limited form of systemic sclerosis, also referred to as the CREST (calcinosis cutis, Raynaud’s phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia) variant (**Chap. 360**), is associated with a chronic course and antientromere antibodies. Mat telangiectasias are an important clue to the diagnosis of this variant as well as the diffuse form of systemic sclerosis because they may be the only cutaneous finding.

Nailfold telangiectasias are pathognomonic signs of the three major autoimmune connective tissue diseases: *lupus erythematosus*, *systemic sclerosis*, and *DM*. They are easily visualized by the naked eye and occur in at least two-thirds of these patients. In both DM and lupus, there is associated nailfold erythema, and in DM, the erythema is often accompanied by “ragged” cuticles and fingertip tenderness. Under 10× magnification or by dermoscopy, the blood vessels in the nailfolds of lupus patients are tortuous and resemble “glomeruli,” whereas in systemic sclerosis and DM, there is a loss of capillary loops and those that remain are markedly dilated.

In *hereditary hemorrhagic telangiectasia* (Osler-Rendu-Weber disease), the lesions usually appear during adolescence (mucosal) and adulthood (cutaneous) and are most commonly seen on the mucous membranes (nasal, orolabial), face, and distal extremities, including under the nails. They represent arteriovenous (AV) malformations of the dermal microvasculature, are dark red in color, and are usually slightly elevated. When the skin is stretched over an individual lesion, an eccentric punctum with radiating legs is seen. Although the degree of systemic involvement varies in this autosomal dominant disease (due primarily to mutations in either the endoglin or activin receptor-like kinase gene), the major symptoms are recurrent epistaxis and gastrointestinal bleeding. The fact that these mucosal telangiectasias are actually AV communications helps to explain their tendency to bleed.

HYPOPIGMENTATION

(**Table 58-9**) Disorders of hypopigmentation are often classified as either diffuse or localized. The classic example of *diffuse hypopigmentation* is *oculocutaneous albinism* (OCA). The most common forms are due to mutations in the tyrosinase gene (type I) or the *P* gene (type II); patients with type IA OCA have a total lack of enzyme activity. At birth, different forms of OCA can appear similar—white hair, gray-blue eyes, and pink-white skin. However, the patients with no tyrosinase activity maintain this phenotype, whereas those with decreased activity will acquire some pigmentation of the eyes, hair, and skin as they age.

TABLE 58-5 Nonscarring Alopecia (Primary Cutaneous Disorders)

	CLINICAL CHARACTERISTICS	PATHOGENESIS	TREATMENT
Telogen effluvium	Diffuse shedding of normal hairs Follows major stress (high fever, severe infection) or change in hormone levels (postpartum) Reversible without treatment	Stress causes more of the asynchronous growth cycles of individual hairs to become synchronous; therefore, larger numbers of growing (anagen) hairs simultaneously enter the dying (telogen) phase	Observation; discontinue any drugs that have alopecia as a side effect; must exclude underlying metabolic causes, e.g., hypothyroidism, hyperthyroidism
Androgenetic alopecia (male pattern; female pattern)	Miniaturization of hairs along the midline of the scalp Recession of the anterior scalp line in men and some women	Increased sensitivity of affected hairs to the effects of androgens—most common Increased levels of circulating androgens (ovarian or adrenal source in women)—less common	If no evidence of hyperandrogenemia, then topical minoxidil; finasteride ^a ; spironolactone (women); hair transplant; low-dose oral minoxidil
Alopecia areata	Well-circumscribed, circular areas of hair loss, 2–5 cm in diameter In extensive cases, coalescence of lesions and/or involvement of other hair-bearing surfaces of the body Pitting or sandpapered appearance of the nails	The germinative zones of the hair follicles are surrounded by T lymphocytes Occasional associated diseases: hyperthyroidism, hypothyroidism, vitiligo, Down syndrome	Topical anthralin or tazarotene; intralesional glucocorticoids; topical contact sensitizers; JAK inhibitors
Tinea capitis	Varies from scaling with minimal hair loss to discrete patches with “black dots” (sites of broken infected hairs) to boggy plaque with pustules (kerion) ^b	Invasion of hairs by dermatophytes, most commonly <i>Trichophyton tonsurans</i>	Oral griseofulvin or terbinafine plus 2.5% selenium sulfide or ketoconazole shampoo; examine family members
Traumatic alopecia ^c	Broken hairs, often of varying lengths Irregular outline in trichotillomania and traction alopecia Fringe sign in traction alopecia	Traction with curlers, rubber bands, tight braiding Exposure to heat or chemicals (e.g., hair straighteners) Mechanical pulling (trichotillomania)	Discontinuation of offending hair style or chemical treatments; diagnosis of trichotillomania may require observation of shaved hairs (for growth) or biopsy, possibly followed by psychotherapy

^aTo date, Food and Drug Administration-approved for men. ^bScarring alopecia can occur at sites of kerions. ^cMay also be scarring, especially late-stage traction alopecia.

The degree of pigment formation is also a function of racial background, and the pigmentary dilution is more readily apparent when patients are compared to their first-degree relatives. The ocular findings in OCA correlate with the degree of hypopigmentation and include decreased visual acuity, nystagmus, photophobia, strabismus, and a lack of normal binocular vision.

TABLE 58-6 Causes of Figurate Skin Lesions

- I. Primary cutaneous disorders
 - A. Tinea
 - B. Urticaria (primary in ≥90% of patients)
 - C. Granuloma annulare
 - D. Erythema annulare centrifugum
 - E. Psoriasis, annular pustular psoriasis
 - F. Interstitial granulomatous drug reaction
- II. Systemic diseases
 - A. Migratory
 - 1. Erythema migrans (CDC case definition is ≥5 cm in diameter)
 - 2. Urticaria (<10% of patients)
 - 3. Erythema gyratum repens
 - 4. Erythema marginatum
 - 5. Pustular psoriasis (generalized and annular forms)
 - 6. Necrolytic migratory erythema (glucagonoma syndrome)^a
 - B. Nonmigratory (may slowly expand)
 - 1. Subacute cutaneous LE, LE tumidus
 - 2. Sarcoidosis
 - 3. Leprosy (borderline, tuberculoid)
 - 4. Secondary syphilis (especially the face)
 - 5. Cutaneous T-cell lymphoma (especially mycosis fungoides)
 - 6. Interstitial granulomatous dermatitis^b
 - 7. Annular erythema of Sjögren's syndrome

^aMigratory erythema with erosions; favors lower extremities and girdle area.

^bUnderlying diseases include rheumatoid arthritis, LE, and granulomatosis with polyangiitis.

Abbreviations: CDC, Centers for Disease Control and Prevention; LE, lupus erythematosus.

The differential diagnosis of *localized hypomelanosis* includes the following primary cutaneous disorders: *postinflammatory hypopigmentation*, *idiopathic guttate hypomelanosis*, *pityriasis (tinea) versicolor*, *vitiligo*, *chemical- or drug-induced leukoderma*, *nevus depigmentosus* (see below), *progressive macular hypomelanosis*, and *piebaldism* (**Table 58-10**). In this group of diseases, the areas of involvement are macules or patches with a decrease or absence of pigmentation. Patients with vitiligo also have an increased incidence of several autoimmune disorders, including Hashimoto's thyroiditis, Graves' disease, pernicious anemia, Addison's disease, uveitis, alopecia areata, chronic mucocutaneous candidiasis, and the autoimmune polyendocrine syndromes (types I and II). Diseases of the thyroid gland are the most frequently associated disorders, occurring in up to 30% of patients with vitiligo. Circulating autoantibodies are often found, and the most common ones are antithyroglobulin, antimicrosomal, and antithyroid-stimulating hormone receptor antibodies.

There are four systemic diseases that should be considered in a patient with skin findings suggestive of vitiligo—*systemic sclerosis*, *melanoma-associated leukoderma*, *onchocerciasis*, and *Vogt-Koyanagi-Harada syndrome*. The vitiligo-like leukoderma seen in patients with

TABLE 58-7 Causes of Acneiform Eruptions

- I. Primary cutaneous disorders
 - A. Acne vulgaris
 - B. Acne rosacea
- II. Drugs, e.g., anabolic steroids, glucocorticoids, lithium, EGFR inhibitors, HER2 inhibitors, MEK inhibitors, iodides
- III. Systemic diseases
 - A. Increased androgen production
 - 1. Adrenal origin, e.g., Cushing's disease, 21-hydroxylase deficiency
 - 2. Ovarian origin, e.g., polycystic ovary syndrome, ovarian hyperthecosis
 - B. Cryptococcosis, disseminated
 - C. Dimorphic fungal infections
 - D. Behçet's disease

Abbreviations: EGFR, epidermal growth factor receptor; HER2, human epidermal growth factor receptor 2; MEK, MAP (mitogen activated protein) kinase.

TABLE 58-8 Causes of Telangiectasias

I.	Primary cutaneous disorders
A.	Linear/branching
1.	Acne rosacea (face)
2.	Actinically damaged skin (face, neck, V of chest)
3.	Venous hypertension (legs)
4.	Generalized essential telangiectasia
5.	Cutaneous collagenous vasculopathy
6.	Within basal cell carcinomas or cutaneous lymphoma
B.	Poikiloderma
1.	Ionizing radiation ^a
C.	Spider angioma
1.	Idiopathic
2.	Pregnancy
II.	Systemic diseases
A.	Linear/branching
1.	Carcinoid (head, neck, upper trunk)
2.	Ataxia-telangiectasia (bulbar conjunctivae, head and neck)
3.	Mastocytosis (within lesions)
B.	Poikiloderma
1.	Dermatomyositis, lupus erythematosus
2.	Mycosis fungoïdes, patch stage
3.	Genodermatoses, e.g., xeroderma pigmentosum, Kindler syndrome
C.	Mat
1.	Systemic sclerosis (scleroderma)
D.	Nailfold
1.	Lupus erythematosus
2.	Systemic sclerosis (scleroderma)
3.	Dermatomyositis
4.	Hereditary hemorrhagic telangiectasia
E.	Papular
1.	Hereditary hemorrhagic telangiectasia
F.	Spider angioma
1.	Cirrhosis ^b

^aBecoming less common. ^bDue to hyperestrogenic state.

systemic sclerosis has a clinical resemblance to idiopathic vitiligo that has begun to repigment as a result of treatment; that is, perifollicular macules of normal pigmentation are seen within areas of depigmentation. The basis of this leukoderma is unknown; there is no evidence of inflammation in areas of involvement, but it can resolve if the underlying connective tissue disease becomes inactive. In contrast to idiopathic vitiligo, melanoma-associated vitiligo-like leukoderma often begins on the trunk, and its appearance, if spontaneous, should prompt a search for metastatic disease. It is also seen in patients undergoing immunotherapy for melanoma, including immune checkpoint-blocking antibodies, with cytotoxic T lymphocytes presumably recognizing cell surface antigens common to melanoma cells and melanocytes, and is associated with a greater likelihood of a clinical response. A history of aseptic meningitis, nontraumatic uveitis, tinnitus, hearing loss, and/or dysacusis points to the diagnosis of the Vogt-Koyanagi-Harada syndrome. In these patients, the face and scalp are the most common locations of pigment loss.

There are two systemic disorders (neurocristopathies) that may have the cutaneous findings of piebaldism (Table 58-9). They are *Shah-Waardenburg syndrome* and *Waardenburg syndrome*. A possible explanation for both disorders is an abnormal embryonic migration or survival of two neural crest-derived elements, one of them being melanocytes and the other myenteric ganglion cells (leading to Hirschsprung disease in Shah-Waardenburg syndrome) or auditory nerve cells (Waardenburg syndrome). The latter syndrome is characterized by congenital sensorineural hearing loss, dystopia canthorum (lateral displacement of the inner canthi but normal interpupillary distance), heterochromic irises, and a broad nasal root, in addition to the piebaldism. The facial

TABLE 58-9 Causes of Hypopigmentation

I.	Primary cutaneous disorders
A.	Diffuse
1.	Generalized vitiligo ^a
B.	Localized
1.	Postinflammatory
2.	Idiopathic guttate hypomelanosis
3.	Pityriasis (tinea) versicolor
4.	Vitiligo ^a
5.	Chemical- or drug-induced leukoderma, e.g., topical imiquimod, oral imatinib
6.	Nevus depigmentosus and pigmentary mosaicism
7.	Progressive macular hypomelanosis
8.	Piebaldism ^a
II.	Systemic diseases
A.	Diffuse
1.	Oculocutaneous albinism ^b
2.	Hermansky-Pudlak syndrome ^{b,c}
3.	Chédiak-Higashi syndrome ^{b,d}
4.	Phenylketonuria
B.	Localized
1.	Systemic sclerosis (scleroderma) ^e
2.	Melanoma-associated vitiligo-like leukoderma, immunotherapy-induced or spontaneous ^e
3.	Sarcoidosis
4.	Cutaneous T-cell lymphoma (especially mycosis fungoïdes)
5.	Tuberculoid and indeterminate leprosy
6.	Onchocerciasis ^e
7.	Linear nevoid hypopigmentation (pigmentary mosaicism) ^{b,f}
8.	Incontinentia pigmenti (stage IV)
9.	Tuberous sclerosis
10.	Waardenburg syndrome and Shah-Waardenburg syndrome
11.	Vogt-Koyanagi-Harada syndrome ^e

^aAbsence of melanocytes in areas of leukoderma; congenital in piebaldism.

^bNormal number of melanocytes. ^cPlatelet storage defect and restrictive lung disease secondary to deposits of ceroid-like material or immunodeficiency; due to mutations in β or δ subunit of adaptor-related protein complex 3 as well as subunits of biogenesis of lysosome-related organelles complex (BLOC)-1, -2, and -3. ^dGiant lysosomal granules and recurrent infections. ^eCan resemble vitiligo due to acquired complete loss of pigment. ^fMinority of patients in a nonreferral setting have systemic abnormalities (musculoskeletal, central nervous system, ocular), previously referred to as hypomelanosis of Ito.

dysmorphism can be explained by the neural crest origin of the connective tissues of the head and neck. Patients with Waardenburg syndrome have been shown to have mutations in four genes, including *PAX-3* and *MITF*, all of which encode transcription factors, whereas patients with Hirschsprung disease plus white spotting have mutations in one of three genes—endothelin 3, endothelin B receptor, and *SOX-10*.

In *tuberous sclerosis*, the earliest cutaneous sign is macular hypomelanosis, referred to as an ash leaf spot. These lesions are often present at birth and are usually multiple; however, detection may require Wood's lamp examination, especially in lightly pigmented individuals. The pigment within them is reduced, but not absent. The average size is 1–3 cm, and the common shapes are polygonal and lance-ovate. Examination of the patient for additional cutaneous signs such as multiple angiofibromas of the face (adenoma sebaceum), ungual and intraoral fibromas, fibrous cephalic plaques, and connective tissue nevi (shagreen patches) is recommended. It is important to remember that an ash leaf spot on the scalp will result in a circumscribed patch of lightly pigmented hair. Internal manifestations include seizures, intellectual disability, central nervous system (CNS) and retinal hamartomas, pulmonary lymphangioleiomyomatosis (women), renal angiomyolipomas, and cardiac rhabdomyomas. The latter can be detected in up to 60% of children (<18 years) with tuberous sclerosis by echocardiography.

Nevus depigmentosus is a stable, well-circumscribed hypomelanosis that is present at birth. There is usually a single oval or rectangular

TABLE 58-10 Hypopigmentation (Primary Cutaneous Disorders, Localized)

	CLINICAL CHARACTERISTICS	WOOD'S LAMP EXAMINATION (UV-A; PEAK = 365 NM)	SKIN BIOPSY SPECIMEN	PATHOGENESIS	TREATMENT
Postinflammatory hypopigmentation	Can develop within active lesions, as in subacute cutaneous lupus, or after the lesion fades, as in atopic dermatitis	Depends on particular disease Usually less enhancement than in vitiligo	Type of inflammatory infiltrate depends on specific disease	Block in transfer of melanin from melanocytes to keratinocytes could be secondary to edema or decrease in contact time Destruction of melanocytes if inflammatory cells attack basal layer of epidermis	Treat underlying inflammatory disease
Idiopathic guttate hypomelanosis	Common; acquired; usually 2–4 mm in diameter Shins and extensor forearms	Less enhancement than vitiligo	Abrupt decrease in epidermal melanin content	Possible somatic mutations as a reflection of aging or UV exposure	None
Pityriasis (tinea) versicolor ^a	Common disorder Upper trunk and neck (shawl-like distribution), groin Young adults Macules have fine white scale when scratched	Golden fluorescence	Hyphal forms and budding yeast in stratum corneum	Invasion of stratum corneum by the yeast <i>Malassezia</i> Yeast is lipophilic and produces C ₉ and C ₁₁ dicarboxylic acids, which in vitro inhibit tyrosinase	Selenium sulfide 2.5% shampoo; topical imidazoles; oral triazoles
Vitiligo	Acquired; progressive Symmetric areas of complete pigment loss Periorificial—around mouth, nose, eyes, nipples, umbilicus, anus Other areas—flexor wrists, extensor distal extremities Segmental form is less common—unilateral, dermatomal-like	More apparent Chalk-white	Absence of melanocytes in well-developed lesions Mild inflammation	Autoimmune phenomenon that results in destruction of melanocytes—primarily cellular (circulating skin-homing autoreactive T cells)	Topical glucocorticoids; topical calcineurin inhibitors; UV-B (narrowband); PUVA; JAK inhibitors; transplants, if stable; depigmentation (topical MBEH), if widespread and treatment-resistant
Chemical- or drug-induced leukoderma	Similar appearance to vitiligo Often begins on hands when associated with chemical exposure Satellite lesions in areas not exposed to chemicals	More apparent Chalk-white	Decreased number or absence of melanocytes	Exposure to chemicals that selectively destroy melanocytes, in particular phenols and catechols (germicides; rubber products) or ingestion of drugs such as imatinib Release of cellular antigens and activation of circulating lymphocytes may explain satellite phenomenon Possible inhibition of KIT receptor	Avoid exposure to offending agent, then treat as vitiligo Drug-induced variant may undergo repigmentation when medication is discontinued
Piebaldism	Autosomal dominant Congenital, stable White forelock Areas of amelanosis contain normally pigmented and hyperpigmented macules of various sizes Symmetric involvement of central forehead, ventral trunk, and mid regions of upper and lower extremities	Enhancement of leukoderma and hyperpigmented macules	Amelanotic areas—few to no melanocytes	Defect in migration of melanoblasts from neural crest to involved skin or failure of melanoblasts to survive or differentiate in these areas Mutations within the <i>KIT</i> protooncogene that encodes the tyrosine kinase receptor for stem cell growth factor (kit ligand)	None; occasionally transplants

^aIf potassium hydroxide (KOH) examination of scale is negative, consider the possibility of progressive macular hypomelanosis.

Abbreviations: MBEH, monobenzylether of hydroquinone; PUVA, psoralens plus ultraviolet A irradiation; UV-B, ultraviolet B irradiation.

lesion, but when there are multiple lesions, the possibility of tuberous sclerosis needs to be considered. In *linear nevoid hypopigmentation*, a term that is replacing hypomelanosis of Ito and segmental or systematized nevus depigmentosus, streaks and swirls of hypopigmentation are observed. Up to one-third of patients in a tertiary care setting had associated abnormalities involving the musculoskeletal system (asymmetry), the CNS (seizures and intellectual disability), and the eyes (strabismus and hypertelorism). Chromosomal mosaicism has

been detected in these patients, lending support to the hypothesis that the cutaneous pattern is the result of the migration of two clones of primordial melanocytes, each with a different pigment potential.

Localized areas of decreased pigmentation are commonly seen as a result of cutaneous inflammation (Table 58-10) and have been observed in the skin overlying active lesions of sarcoidosis (see “Papulonodular Skin Lesions,” below) as well as in CTCL. Cutaneous infections also present as disorders of hypopigmentation, and in *tuberculoid*

leprosy, there are a few asymmetric patches of hypomelanosis that have associated anesthesia, anhidrosis, and alopecia. Biopsy specimens of the palpable border show dermal granulomas that contain rare, if any, *Mycobacterium leprae* organisms.

HYPERPIGMENTATION

(**Table 58-11**) Disorders of hyperpigmentation are also divided into two major groups—localized and diffuse. The localized forms are due to an epidermal alteration, a proliferation of melanocytes, or an increase in pigment production. Both acanthosis nigricans and seborrheic keratoses belong to the first group. *Acanthosis nigricans* can be a reflection of an internal malignancy, most commonly of the gastrointestinal tract, and it appears as velvety hyperpigmentation, primarily in flexural areas. However, in the majority of patients, acanthosis nigricans is associated with obesity and insulin resistance, although it may be a reflection of an endocrinopathy such as acromegaly, Cushing's syndrome, polycystic ovary syndrome, or insulin-resistant diabetes mellitus (type A, type B, and lipodystrophic forms). *Seborrheic keratoses* are common lesions, but in one rare clinical setting, they are a sign of systemic disease, and that setting is the sudden appearance of multiple lesions, often with an inflammatory base and in association with acrochordons (skin tags) and acanthosis nigricans. This is termed the *sign of Leser-Trélat* and alerts the clinician to search for an internal malignancy.

A proliferation of melanocytes results in the following pigmented lesions: *lentigo*, *melanocytic nevus*, and *melanoma* (**Chap. 76**). In an adult, the majority of lentigines are related to sun exposure, which explains their distribution. However, in the Peutz-Jeghers and LEOPARD (lentigines; ECG abnormalities, primarily conduction defects; ocular hypertelorism; pulmonary stenosis and subaortic valvular stenosis; abnormal genitalia [cryptorchidism, hypospadias]; retardation of growth; and deafness [sensorineural]) syndromes, lentigines do serve as a clue to systemic disease. In *LEOPARD/Noonan with multiple lentigines syndrome*, hundreds of lentigines develop during childhood and are scattered over the entire surface of the body. The lentigines in patients with *Peutz-Jeghers syndrome* are located primarily around the nose and mouth, on the hands and feet, and within the oral cavity. While the pigmented macules on the face may fade with age, the oral lesions persist. However, similar intraoral lesions are also seen in Addison's disease, in Laugier-Hunziker syndrome (no internal manifestations), and as a normal finding in darkly pigmented individuals. Patients with this autosomal dominant syndrome (due to mutations in a novel serine threonine kinase gene) have multiple benign polyps of the gastrointestinal tract, testicular or ovarian tumors, and an increased risk of developing gastrointestinal (primarily colon) and pancreatic cancers.

In the *Carney complex*, numerous lentigines are also seen, but they are in association with cardiac myxomas. This autosomal dominant disorder is also known as the *LAMB* (lentigines, atrial myxomas, mucocutaneous myxomas, and blue nevi) *syndrome* or *NAME* (nevus, atrial myxoma, myxoid neurofibroma, and ephelides [freckles]) *syndrome*. These patients can also have evidence of endocrine overactivity in the form of Cushing's syndrome (pigmented nodular adrenocortical disease) and acromegaly.

The third type of localized hyperpigmentation is due to a local increase in pigment production, and it includes *ephelides* and *café au lait macules* (CALMs). While a single CALM can be seen in up to 10% of the normal population, the presence of multiple or large-sized CALMs raises the possibility of an associated genodermatosis, for example, neurofibromatosis (NF) or McCune-Albright syndrome. CALMs are flat, uniformly brown in color (usually two shades darker than uninvolved skin), and can vary in size from 0.5 to 12+ cm. More than 90% of adult patients with *type I NF* will have six or more CALMs measuring ≥1.5 cm in diameter. Additional findings are discussed in the section on neurofibromas (see "Papulonodular Skin Lesions," below). In comparison with NF, the CALMs in patients with *McCune-Albright syndrome* (polyostotic fibrous dysplasia with precocious puberty in females due to mosaicism for an activating mutation

TABLE 58-11 Causes of Hyperpigmentation

- I. Primary cutaneous disorders
 - A. Localized
 - 1. Epidermal alteration
 - a. Seborrheic keratosis
 - b. Pigmented actinic keratosis
 - 2. Proliferation of melanocytes
 - a. Lentigo
 - b. Melanocytic nevus (mole)
 - c. Melanoma
 - 3. Increased pigment production
 - a. Ephelide (freckle)
 - b. Café au lait macule
 - c. Postinflammatory hyperpigmentation (also dermal)
 - d. Melasma (also dermal)
 - 4. Dermal pigmentation
 - a. Fixed drug eruption
 - B. Localized and diffuse
 - 1. Drugs (e.g., minocycline, hydroxychloroquine, bleomycin)
- II. Systemic diseases
 - A. Localized
 - 1. Epidermal alteration
 - a. Acanthosis nigricans (insulin resistance > other endocrine disorders, paraneoplastic)
 - b. Seborrheic keratoses (sign of Leser-Trélat)
 - 2. Proliferation of melanocytes
 - a. Lentigines (Peutz-Jeghers and LEOPARD/Noonan with multiple lentigines syndromes; xeroderma pigmentosum)
 - b. Melanocytic nevi (Carney complex [LAMB and NAME syndromes])^a
 - 3. Increased pigment production
 - a. Café au lait macules (neurofibromatosis, Legius syndrome, McCune-Albright syndrome^b)
 - b. Urticaria pigmentosa^c
 - 4. Dermal pigmentation
 - a. Incontinentia pigmenti (stage III)
 - b. Dyskeratosis congenita
 - 5. Dermal deposits
 - a. Exogenous ochronosis
 - b. Localized argyria
 - B. Diffuse
 - 1. Endocrinopathies
 - a. Addison's disease
 - b. Nelson syndrome
 - c. Ectopic ACTH syndrome
 - d. Hyperthyroidism
 - 2. Metabolic
 - a. Porphyria cutanea tarda
 - b. Hemochromatosis
 - c. Vitamin B₁₂, folate deficiency
 - d. Pellagra
 - e. Malabsorption, including Whipple's disease
 - 3. Melanosis secondary to metastatic melanoma
 - 4. Autoimmune
 - a. Primary biliary cholangitis
 - b. Systemic sclerosis (scleroderma)
 - c. POEMS syndrome
 - d. Eosinophilia-myalgia syndrome^d
 - 5. Drugs (e.g., cyclophosphamide) and metals (e.g., silver)

^aAlso lentigines. ^bPolyostotic fibrous dysplasia. ^cSee also "Papulonodular Skin Lesions." ^dLate 1980s.

Abbreviations: LAMB, lentigines, atrial myxomas, mucocutaneous myxomas, and blue nevi; LEOPARD, lentigines, ECG abnormalities, ocular hypertelorism, pulmonary stenosis and subaortic valvular stenosis, abnormal genitalia, retardation of growth, and deafness (sensorineural); NAME, nevi, atrial myxoma, myxoid neurofibroma, and ephelides (freckles); POEMS, polyneuropathy, organomegaly, endocrinopathies, M-protein, and skin changes.

in a G protein [G_a] gene) are usually larger, are more irregular in outline, and tend to respect the midline.

In *incontinentia pigmenti*, dyskeratosis congenita, and bleomycin pigmentation, the areas of localized hyperpigmentation form a pattern—swirls and streaks in the first, reticulated in the second, and flagellate in the third. In *dyskeratosis congenita*, atrophic reticulated hyperpigmentation is seen on the neck, trunk, and thighs and is accompanied by nail dystrophy, pancytopenia, and leukoplakia of the oral and anal mucosae. The latter often develops into squamous cell carcinoma. In addition to the flagellate pigmentation (linear streaks) on the trunk, patients receiving bleomycin often have hyperpigmentation overlying the elbows, knees, and small joints of the hand.

Localized hyperpigmentation is seen as a side effect of several other *systemic medications*, including those that produce fixed drug reactions (nonsteroidal anti-inflammatory drugs [NSAIDs], sulfonamides, barbiturates, and tetracyclines) and those that can complex with melanin or iron (antimalarials and minocycline). Fixed drug eruptions recur in the exact same location as circular areas of erythema that can become bullous and then resolve as brown macules. The eruption usually appears within hours of readministration of the offending agent, and common locations include the genitalia, distal extremities, and perioral region. Chloroquine and hydroxychloroquine produce gray-brown to blue-black discoloration of the shins, hard palate, and face, while blue macules (often misdiagnosed as bruises) can be seen on the lower extremities and in sites of inflammation with prolonged minocycline administration. Estrogen in oral contraceptives can induce melasma—symmetric brown patches on the face, especially the cheeks, upper lip, and forehead. Similar changes are seen in pregnancy and in patients receiving phenytoin.

In the diffuse forms of hyperpigmentation, the darkening of the skin may be of equal intensity over the entire body or may be accentuated in sun-exposed areas. The causes of diffuse hyperpigmentation can be divided into four major groups—endocrine, metabolic, autoimmune, and drugs. The endocrinopathies that frequently have associated hyperpigmentation include *Addison's disease*, *Nelson's syndrome*, and *ectopic adrenocorticotrophic hormone (ACTH) syndrome*. In these diseases, the increased pigmentation is diffuse but is accentuated in sun-exposed areas, as well as in the palmar creases, sites of friction, and scars. An overproduction of the pituitary hormones α-MSH (melanocyte-stimulating hormone) and ACTH can lead to an increase in melanocyte activity. These peptides are products of the proopiomelanocortin gene and exhibit homology; for example, α-MSH and ACTH share 13 amino acids. A minority of patients with Cushing's disease or hyperthyroidism have generalized hyperpigmentation.

The metabolic causes of hyperpigmentation include *porphyria cutanea tarda* (PCT), *hemochromatosis*, *vitamin B₁₂ deficiency*, *folic acid deficiency*, *pellagra*, and *malabsorption*, including *Whipple's disease*. In patients with PCT (see “Vesicles/Bullae,” below), the skin darkening is seen in sun-exposed areas and is a reflection of the photoreactive properties of porphyrins. The increased level of iron in the skin of patients with type 1 hemochromatosis stimulates melanin pigment production and leads to the classic bronze color. Patients with pellagra have a brown discoloration of the skin, especially in sun-exposed areas, as a result of nicotinic acid (niacin) deficiency. In the areas of increased pigmentation, there is a thin, varnish-like scale. These changes are also seen in patients who are vitamin B₆ deficient, have functioning carcinoid tumors (increased consumption of niacin), or take isoniazid. Approximately 50% of the patients with Whipple's disease have an associated generalized hyperpigmentation in association with diarrhea, weight loss, arthritis, and lymphadenopathy. A diffuse, slate-blue to gray-brown color is seen in patients with *melanosis secondary to metastatic melanoma*. The color reflects widespread deposition of melanin within the dermis as a result of the high concentration of circulating melanin precursors.

Of the autoimmune diseases associated with diffuse hyperpigmentation, *primary biliary cholangitis* and *systemic sclerosis* are the most common, and occasionally, both disorders are seen in the same patient. The skin is dark brown in color, especially in sun-exposed areas. In primary biliary cholangitis, the hyperpigmentation is accompanied by

pruritus, jaundice, and xanthomas, whereas in systemic sclerosis, it is accompanied by sclerosis of the extremities, face, and, less commonly, the trunk. Additional clues to the diagnosis of systemic sclerosis are mat and cuticular telangiectasias, calcinosis cutis, Raynaud's phenomenon, and distal ulcerations (see “Telangiectasias,” above). The differential diagnosis of cutaneous sclerosis with hyperpigmentation includes POEMS (*polyneuropathy; organomegaly [liver, spleen, lymph nodes]; endocrinopathies [impotence, gynecomastia]; M-protein; and skin changes*) syndrome. The skin changes include hyperpigmentation, induration, hypertrichosis, angiomas, clubbing, and facial lipoatrophy.

Diffuse hyperpigmentation that is due to drugs or metals can result from one of several mechanisms—induction of melanin pigment formation, complexing of the drug or its metabolites to melanin, and deposits of the drug in the dermis. Busulfan, cyclophosphamide, 5-fluorouracil, and inorganic arsenic induce pigment production. Complexes containing melanin or iron plus the drug or its metabolites are seen in patients receiving minocycline, and a diffuse, brown-gray, muddy appearance within sun-exposed areas may develop, in addition to pigmentation of the mucous membranes, teeth, nails, bones, and thyroid. Administration of amiodarone can result in both a phototoxic eruption (exaggerated sunburn) and/or a slate-gray to violaceous discoloration of sun-exposed skin. Biopsy specimens of the latter show yellow-brown granules in dermal macrophages, which represent intralysosomal accumulations of lipids, amiodarone, and its metabolites. Actual deposits of a particular drug or metal in the skin are seen with silver (argyria), where the skin appears blue-gray in color; gold (chrysiasis), where the skin has a brown to blue-gray color; and clofazimine, where the skin appears reddish brown. The associated pigmentation is accentuated in sun-exposed areas, and discoloration of the eye is seen with gold (sclerae) and clofazimine (conjunctivae).

VESICLES/BULLAE

(**Table 58-12**) Depending on their size, cutaneous blisters are referred to as *vesicles* (<1 cm) or *bullae* (>1 cm). The primary autoimmune blistering disorders include *pemphigus vulgaris*, *pemphigus foliaceus*, *paraneoplastic pemphigus*, *bullosic pemphigoid*, *gestational pemphigoid*, *cicatricial pemphigoid*, *epidermolysis bullosa acquista*, *linear IgA bullous dermatosis (LABD)*, and *dermatitis herpetiformis (Chap. 59)*.

Vesicles and bullae are also seen in *contact dermatitis*, both allergic and irritant forms (**Chap. 57**). When there is a linear arrangement of vesicular lesions, an exogenous cause or herpes zoster should be suspected. Bullous disease secondary to the ingestion of drugs can take one of several forms, including phototoxic eruptions, isolated bullae, Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN) (**Chap. 60**). Clinically, phototoxic eruptions resemble an exaggerated sunburn with diffuse erythema and bullae in sun-exposed areas. The most commonly associated drugs are doxycycline, quinolones, voriconazole, thiazides, NSAIDs, vemurafenib, and psoralens. The development of a phototoxic eruption is dependent on the doses of both the drug and ultraviolet (UV)-A irradiation.

Toxic epidermal necrolysis is characterized by bullae that arise on widespread areas of tender erythema and then slough. This results in large areas of denuded skin. The associated morbidity, such as sepsis, and mortality rates are relatively high and are a function of the extent of epidermal necrosis. In addition, these patients may also have involvement of the mucous membranes and respiratory and intestinal tracts. Drugs are the primary cause of TEN, and the most common offenders are aromatic anticonvulsants (phenytoin, barbiturates, carbamazepine), sulfonamides, aminopenicillins, allopurinol, and NSAIDs. Generalized bullous fixed drug eruption, severe acute graft-versus-host disease (grade 4), vancomycin-induced LABD, and flares of cutaneous lupus can also resemble TEN.

In *erythema multiforme (EM)*, the primary lesions are pink-red macules and edematous papules, the centers of which may become vesicular. In contrast to a morbilliform exanthem, the clue to the diagnosis of EM, and especially SJS, is the development of a “dusky” violet color in the center of the lesions. Target lesions are also characteristic of EM and arise as a result of active centers and borders in combination

TABLE 58-12 Causes of Vesicles/Bullae

I. Primary mucocutaneous diseases
A. Primary blistering diseases (autoimmune)
1. Pemphigus, foliaceus and vulgaris ^a
2. Bullous pemphigoid ^b
3. Gestational pemphigoid ^b
4. Cicatricial pemphigoid ^b
5. Dermatitis herpetiformis ^{b,c}
6. Linear IgA bullous dermatosis ^b
7. Epidermolysis bullosa acquisita ^{b,d}
B. Secondary blistering diseases
1. Contact dermatitis ^{a,b}
2. Erythema multiforme ^e
3. Stevens-Johnson syndrome ^e
4. Toxic epidermal necrolysis ^e
5. Bullous fixed drug eruption, including generalized variant ^e
6. Pseudoporphyria, drug- or tanning booth-induced
C. Infections
1. Varicella-zoster virus ^{a,f}
2. Herpes simplex virus ^{a,f}
3. Enteroviruses, e.g., hand-foot-and-mouth disease ^f
4. SARS-CoV-2
5. Staphylococcal scalded-skin syndrome ^{a,g}
6. Bullous impetigo ^a
7. Bullous tinea
II. Systemic diseases
A. Autoimmune
1. Paraneoplastic pemphigus ^a (bronchiolitis obliterans)
B. Infections
1. Cutaneous emboli ^b
C. Metabolic
1. Diabetic bullae ^{a,b}
2. Porphyria cutanea tarda ^b
3. Porphyria variegata ^b
4. Bullous dermatosis of hemodialysis ^b (less often associated with peritoneal dialysis and also referred to as pseudoporphyria)
D. Ischemia
1. Coma bullae
E. Secondary blistering diseases
1. Toxic epidermal necrolysis ^e (respiratory and gastrointestinal tracts can be involved)

^aIntraepidermal. ^bSubepidermal. ^cAssociated with gluten enteropathy. ^dAssociated with inflammatory bowel disease. ^eDegeneration of cells within the basal layer of the epidermis can give impression split is subepidermal. ^fAlso systemic. ^gIn adults, associated with renal failure and immunocompromised state.

with centrifugal spread. However, target lesions need not be present to make the diagnosis of EM.

EM has been subdivided into two major groups: (1) EM minor due to herpes simplex virus (HSV); and (2) EM major due to HSV, *Mycoplasma pneumonia*, or, occasionally, other viruses, *Chlamydia*, or drugs. Involvement of the mucous membranes (ocular, nasal, oral, and genital) is seen more commonly in the latter form, and in patients with *Mycoplasma pneumoniae*-induced rash and mucositis (MIRM), there may be minimal cutaneous involvement. Hemorrhagic crusts of the lips are characteristic of EM major and SJS as well as herpes simplex, pemphigus vulgaris, and paraneoplastic pemphigus. Fever, malaise, myalgias, sore throat, and cough may precede or accompany the eruption. The lesions of EM usually resolve over 2–4 weeks but may be recurrent, especially when due to HSV. In addition to HSV (in which lesions usually appear 7–12 days after the viral eruption), EM can also follow vaccinations, radiation therapy, and exposure to environmental toxins, including the oleoresin in poison ivy.

Induction of SJS is most often due to drugs, especially sulfonamides, aromatic anticonvulsants, lamotrigine, aminopenicillins,

and nonnucleoside reverse transcriptase inhibitors (e.g., nevirapine). Widespread dusky macules and significant mucosal involvement are characteristic of SJS, and the cutaneous lesions may or may not develop epidermal detachment. If the latter occurs, by definition, it is limited to <10% of the body surface area (BSA). Greater involvement leads to the diagnosis of SJS/TEN overlap (10–30% BSA) or TEN (>30% BSA).

In addition to primary blistering disorders and hypersensitivity reactions, bacterial and viral infections can lead to vesicles and bullae. The most common infectious agents are HSV (Chap. 192), varicella-zoster virus (Chap. 193), and *S. aureus* (Chap. 147).

Staphylococcal scalded-skin syndrome (SSSS) and *bullous impetigo* are two blistering disorders associated with staphylococcal (phage group II) infection. In SSSS, the initial findings are redness and tenderness of the central face, neck, trunk, and intertriginous zones. This is followed by short-lived flaccid bullae and a slough or exfoliation of the superficial epidermis. Crusted areas then develop, characteristically around the mouth in a radial pattern. SSSS is distinguished from TEN by the following features: younger age group (primarily infants and toddlers), more superficial site of blister formation, no oral lesions, shorter course, lower morbidity and mortality rates, and an association with staphylococcal exfoliative toxin (“exfoliatin”), not drugs. A rapid diagnosis of SSSS versus TEN can be made by a frozen section of the blister roof or exfoliative cytology of the blister contents. In SSSS, the site of staphylococcal infection is usually extracutaneous (conjunctivitis, rhinorrhea, otitis media, pharyngitis, tonsillitis), and the cutaneous lesions are sterile, whereas in bullous impetigo, the skin lesions are the site of infection. Impetigo is more localized than SSSS and usually presents with honey-colored crusts. Occasionally, superficial purulent blisters also form. *Cutaneous emboli* from gram-negative infections may present as isolated bullae, but the base of the lesion is purpuric or necrotic, and it may develop into an ulcer (see “Purpura,” below).

Several metabolic disorders are associated with blister formation, including diabetes mellitus, renal failure, and porphyria. Local hypoxemia secondary to decreased cutaneous blood flow can also produce blisters, which explains the presence of bullae over pressure points in comatose patients (coma bullae). In *diabetes mellitus*, tense bullae with clear sterile viscous fluid arise on normal skin. The lesions can be as large as 6 cm in diameter and are located on the distal extremities. There are several types of porphyria, but the most common form with cutaneous findings is *porphyria cutanea tarda* (PCT). In sun-exposed areas (primarily the hands), the skin is very fragile, with trauma leading to erosions mixed with tense vesicles. These lesions then heal with scarring and formation of milia; the latter are firm, 1- to 2-mm white or yellow papules that represent epidermoid cysts. Associated findings can include hypertrichosis of the lateral malar region (men) or face (women) and, in sun-exposed areas, hyperpigmentation and firm sclerotic plaques. An elevated level of urinary uroporphyrins confirms the diagnosis and is due to a decrease in uroporphyrinogen decarboxylase activity. PCT can be exacerbated by alcohol, hemochromatosis and other forms of iron overload, chlorinated hydrocarbons, hepatitis C virus and HIV infections, and hepatomas.

The differential diagnosis of PCT includes (1) *porphyria variegata*—the skin signs of PCT plus the systemic findings of acute intermittent porphyria; it has a diagnostic plasma porphyrin fluorescence emission at 626 nm; (2) *drug-induced pseudoporphyria*—the clinical and histologic findings are similar to PCT, but porphyrins are normal; etiologic agents include naproxen and other NSAIDs, furosemide, tetracycline, and voriconazole; (3) *bullous dermatosis of hemodialysis*—the same appearance as PCT, but porphyrins are usually normal or occasionally borderline elevated; patients have chronic renal failure and are on hemodialysis; (4) *PCT associated with hepatomas and hemodialysis*; and (5) *epidermolysis bullosa acquisita* (Chap. 59).

EXANTHEMS

(Table 58-13) Exanthems are characterized by an acute generalized eruption. The most common presentation is erythematous macules and papules (morbilliform) and less often confluent blanching erythema (scarlatiniform). Morbilliform eruptions are usually due to either drugs or viral infections. For example, up to 5% of patients receiving penicillins,

TABLE 58-13 Causes of Exanthms

I. Morbilliform
A. Drugs
B. Viral
1. Rubeola (measles)
2. Rubella
3. Erythema infectiosum (erythema of cheeks; reticulated on extremities)
4. Epstein-Barr virus, echovirus, coxsackievirus, CMV, adenovirus, HHV-6/HHV-7 ^a , dengue, Zika, chikungunya, SARS-CoV-2, and West Nile virus infections
5. HIV seroconversion exanthem (plus mucosal ulcerations)
C. Bacterial
1. Typhoid fever
2. Early secondary syphilis
3. Early <i>Rickettsia</i> infections
4. Early meningococcemia
5. Ehrlichiosis
D. Acute graft-versus-host disease
E. Kawasaki disease
II. Scarlatiniform
A. Scarlet fever
B. Toxic shock syndrome
C. Kawasaki disease
D. Early staphylococcal scalded-skin syndrome

^aPrimary infection in infants and reactivation in the setting of immunosuppression.

Abbreviations: CMV, cytomegalovirus; HHV, human herpesvirus; HIV, human immunodeficiency virus.

sulfonamides, phenytoin, or nevirapine will develop a maculopapular eruption. Accompanying signs may include pruritus, fever, eosinophilia, transaminitis, and transient lymphadenopathy (**Chap. 60**). Similar maculopapular eruptions are seen in the classic childhood viral exanthems, including (1) *rubeola* (measles)—a prodrome of coryza, cough, and conjunctivitis followed by Koplik's spots on the buccal mucosa; the eruption begins behind the ears, at the hairline, and on the forehead and then spreads down the body, often becoming confluent; (2) *rubella*—the eruption begins on the forehead and face and then spreads down the body; it resolves in the same order and is associated with retroauricular and suboccipital lymphadenopathy; and (3) *erythema infectiosum* (fifth disease)—erythema of the cheeks is followed by a reticulated pattern on the extremities; it is secondary to a parvovirus B19 infection, and an associated arthritis is seen in adults.

Both measles and rubella can occur in unvaccinated adults, and an atypical form of measles is seen in adults immunized with either killed measles vaccine or killed vaccine followed in time by live vaccine. In contrast to classic measles, the eruption of atypical measles begins on the palms, soles, wrists, and ankles, and the lesions may become purpuric. The patient with atypical measles can have pulmonary involvement and be quite ill. Rubelliform and roseoliform eruptions are also associated with *Epstein-Barr virus* (5–15% of patients), *echovirus*, *coxsackievirus*, *cytomegalovirus*, *adenovirus*, SARS-CoV-2, and *dengue*, *chikungunya*, and *West Nile virus* infections. Detection of specific IgM antibodies or fourfold elevations in IgG antibodies often allows the proper diagnosis, but polymerase chain reaction (PCR) is gradually replacing serologic assays. Occasionally, a maculopapular drug eruption is a reflection of an underlying viral infection. For example, ~95% of the patients with infectious mononucleosis who are given ampicillin will develop a rash.

Of note, early in the course of infections with *Rickettsia* and meningococcus, prior to the development of petechiae and purpura, the lesions may be erythematous macules and papules. This is also the case in chickenpox prior to the development of vesicles. Maculopapular eruptions are associated with early *HIV* infection, early secondary *syphilis*, *typhoid fever*, and *acute graft-versus-host disease*. In the last, lesions frequently begin on the dorsal hands and forearms; the macular rose spots of typhoid fever involve primarily the anterior trunk.

The prototypic *scarlatiniform* eruption is seen in *scarlet fever* and is due to an erythrogenic toxin produced by bacteriophage-containing group A β-hemolytic streptococci, most commonly in the setting of pharyngitis. This eruption is characterized by diffuse erythema, which begins on the neck and upper trunk, and red follicular puncta. Additional findings include a white strawberry tongue (white coating with red papillae) followed by a red strawberry tongue (red tongue with red papillae); petechiae of the palate; a facial flush with circumoral pallor; linear petechiae in the antecubital fossae; and desquamation of the involved skin, palms, and soles 5–20 days after onset of the eruption. A similar desquamation of the palms and soles is seen in toxic shock syndrome (TSS), in Kawasaki disease, and after severe febrile illnesses. Certain strains of staphylococci also produce an erythrogenic toxin that leads to the same clinical findings as in streptococcal scarlet fever, except that the anti-streptolysin O or DNase B titers are not elevated.

In *toxic shock syndrome*, staphylococcal (phage group I) infections produce an exotoxin (TSST-1) that causes the fever and rash as well as enterotoxins. Initially, the majority of cases were reported in menstruating women who were using tampons. However, other sites of infection, including wounds and nasal packing, can lead to TSS. The diagnosis of TSS is based on clinical criteria (**Chap. 147**), and three of these involve mucocutaneous sites (diffuse erythema of the skin, desquamation of the palms and soles 1–2 weeks after onset of illness, and involvement of the mucous membranes). The latter is characterized as hyperemia of the vagina, oropharynx, or conjunctivae. Similar systemic findings have been described in *streptococcal toxic shock syndrome* (**Chap. 148**), and although an exanthem is seen less often than in TSS due to a staphylococcal infection, the underlying infection is often in the soft tissue (e.g., cellulitis).

The cutaneous eruption in *Kawasaki disease* (**Chap. 363**) is polymorphous, but the two most common forms are morbilliform and scarlatiniform. Additional mucocutaneous findings include bilateral conjunctival injection; erythema and edema of the hands and feet followed by desquamation; and diffuse erythema of the oropharynx, red strawberry tongue, and dry fissured lips. This clinical picture can resemble TSS and scarlet fever, but clues to the diagnosis of Kawasaki disease are cervical lymphadenopathy, cheilitis, and thrombocytosis. The most serious associated systemic finding in this disease is coronary aneurysms secondary to arteritis. Seen primarily in children, SARS-CoV-2-associated multisystem inflammatory syndrome must be distinguished from Kawasaki disease. Scarlatiniform eruptions are also seen in the early phase of SSSS (see “Vesicles/Bullae,” above), in young adults with *Arcanobacterium haemolyticum* infection, and as reactions to drugs.

URTICARIA

(**Table 58-14**) *Urticaria* (hives) are transient lesions that are composed of a central wheal surrounded by an erythematous halo or flare. Individual lesions are round, oval, or figurate and are often pruritic. Acute and chronic urticarias have a wide variety of allergic etiologies

TABLE 58-14 Causes of Urticaria and Angioedema

I. Primary cutaneous disorders
A. Acute and chronic urticaria ^a
B. Physical urticaria
1. Dermographism
2. Solar urticaria ^b
3. Cold urticaria ^b
4. Cholinergic urticaria ^b
C. Angioedema (hereditary and acquired) ^{b,c}
II. Systemic diseases
A. Urticarial vasculitis
B. Hepatitis B or C viral infection, SARS-CoV-2 infection
C. Serum sickness
D. Angioedema (hereditary and acquired)

^aA small minority develop anaphylaxis. ^bAlso systemic. ^cAcquired angioedema can be idiopathic, associated with a lymphoproliferative disorder, or due to a drug, e.g., angiotensin-converting enzyme (ACE) inhibitors.

and reflect edema in the dermis. Urticular lesions can also be seen in patients with mastocytosis (urticaria pigmentosa), hypo- or hyperthyroidism, Schnitzler's syndrome, and systemic-onset juvenile idiopathic arthritis (Still's disease). In both juvenile- and adult-onset Still's disease, the lesions coincide with the fever spike, are transient, and are due to dermal infiltrates of neutrophils; the latter is also referred to as neutrophilic urticarial dermatosis.

The common *physical urticarias* include dermographism, solar urticaria, cold urticaria, and cholinergic urticaria. Patients with *dermographism* exhibit linear wheals following minor pressure or scratching of the skin and may be a contributing factor to pruritic dermatoses. It is a common disorder, affecting ~5% of the population. *Solar urticaria* characteristically occurs within minutes of sun exposure and is a skin sign of one systemic disease—erythropoietic protoporphyrin. In addition to the urticaria, these patients have subtle pitted scarring of the nose and hands. *Cold urticaria* is precipitated by exposure to the cold, and therefore, exposed areas are usually affected. In occasional patients, the disease is associated with abnormal circulating proteins—more commonly cryoglobulins and less commonly cryofibrinogens. Additional systemic symptoms include wheezing and syncope, thus explaining the need for these patients to avoid swimming in cold water. Autosomal dominantly inherited cold urticaria is associated with dysfunction of cryopyrin. *Cholinergic urticaria* is precipitated by heat, exercise, or emotion and is characterized by small wheals with relatively large flares. It is occasionally associated with wheezing.

Whereas urticarias are the result of dermal edema, subcutaneous edema leads to the clinical picture of *angioedema*. Sites of involvement include the eyelids, lips, tongue, larynx, and gastrointestinal tract as well as the subcutaneous tissue. Angioedema occurs alone or in combination with urticaria, including urticarial vasculitis and the physical urticarias. Both acquired and hereditary (autosomal dominant) forms of angioedema occur (*Chap. 354*), and in the latter, urticaria is rarely, if ever, seen.

Urticular vasculitis is an immune complex disease that may be confused with simple urticaria. In contrast to simple urticaria, individual lesions tend to last longer than 24 h and usually develop central petechiae that can be observed even after the urticarial phase has resolved. The patient may also complain of burning rather than pruritus. On biopsy, there is a leukocytoclastic vasculitis of the small dermal blood vessels. Although urticarial vasculitis may be idiopathic in origin, it can be a reflection of an underlying systemic illness such as lupus erythematosus, Sjögren's syndrome, or hereditary complement deficiency. There is a spectrum of urticarial vasculitis that ranges from purely cutaneous to multisystem involvement. The most common systemic signs and symptoms are arthralgias and/or arthritis, nephritis, and crampy abdominal pain, with asthma and chronic obstructive lung disease seen less often. Hypocomplementemia occurs in one- to two-thirds of patients, even in the idiopathic cases. Urticarial vasculitis can also be seen in patients with *hepatitis B* and *hepatitis C* infections and serum sickness, but is usually not seen in serum sickness-like illnesses (e.g., due to cefaclor, minocycline).

PAPULONODULAR SKIN LESIONS

(*Table 58-15*) In the *papulonodular diseases*, the lesions are elevated above the surface of the skin and may coalesce to form larger plaques. The location, consistency, and color of the lesions are the keys to their diagnosis; this section is organized on the basis of color.

WHITE LESIONS

In *calcinosis cutis*, there are firm white to white-yellow papules with an irregular surface. When the contents are expressed, a chalky white material is seen. *Dystrophic calcification* is seen at sites of previous inflammation or damage to the skin. It develops in acne scars as well as on the distal extremities of patients with systemic sclerosis and in the subcutaneous tissue and intermuscular fascial planes in DM. The latter is more extensive and is more commonly seen in children. An elevated calcium phosphate product, most commonly due to secondary hyperparathyroidism in the setting of renal failure, can lead to nodules of *metastatic calcinosis cutis*, which tend to be subcutaneous and

TABLE 58-15 Papulonodular Skin Lesions According to Color Groups

- I. White
 - A. Calcinosis cutis
 - B. Osteoma cutis (also skin-colored or blue)
- II. Skin-colored
 - A. Rheumatoid nodules
 - B. Neurofibromas (von Recklinghausen's disease [NF1])
 - C. Angiofibromas (tuberous sclerosis, MEN syndrome, type 1; also pink-red)
 - D. Neuromas (MEN syndrome, type 2b)
 - E. Adnexal tumors
 - 1. Basal cell carcinomas (basal cell nevus syndrome)
 - 2. Tricholemmomas (Cowden disease)
 - 3. Fibrofolliculomas (Birt-Hogg-Dubé syndrome)
 - F. Osteomas (arise in skull and jaw in Gardner syndrome)
 - G. Primary cutaneous disorders
 - 1. Epidermal inclusion cysts^a
 - 2. Lipomas
- III. Pink/translucent^b
 - A. Amyloidosis, primary systemic
 - B. Papular mucinosis/scleromyxedema
 - C. Multicentric reticulohistiocytosis
- IV. Yellow
 - A. Xanthomas
 - B. Tophi
 - C. Necrobiosis lipoidica
 - D. Pseudoxanthoma elasticum
 - E. Sebaceous adenomas (Muir-Torre syndrome)
- V. Red^b
 - A. Papules
 - 1. Angiokeratomas (Fabry disease and related lysosomal storage diseases)^c
 - 2. Bacillary angiomatosis (primarily in AIDS)
 - B. Papules/plaques
 - 1. Cutaneous lupus erythematosus
 - 2. Lymphoma cutis
 - 3. Leukemia cutis
 - 4. Sweet syndrome
 - C. Nodules
 - 1. Panniculitis
 - 2. Medium-sized vessel vasculitis (e.g., cutaneous polyarteritis nodosa)
 - D. Primary cutaneous disorders
 - 1. Arthropod bites
 - 2. Cherry hemangiomas
 - 3. Infections, e.g., streptococcal cellulitis, sporotrichosis
 - 4. Polymorphous light eruption
 - 5. Cutaneous lymphoid hyperplasia (lymphocytoma cutis, pseudolymphoma)
- VI. Red-brown^b
 - A. Sarcoidosis
 - B. Urticaria pigmentosa
 - C. Erythema elevatum diutinum (chronic leukocytoclastic vasculitis)
 - D. Lupus vulgaris
- VII. Blue^b
 - A. Venous malformations (e.g., blue rubber bleb syndrome)
 - B. Primary cutaneous disorders
 - 1. Venous lake
 - 2. Blue nevus
- VIII. Violaceous
 - A. Lupus pernio (sarcoidosis)
 - B. Lymphoma cutis
 - C. Cutaneous lupus erythematosus
- IX. Purple
 - A. Kaposi's sarcoma, acral angiokeratoma (pseudo-Kaposi's sarcoma)
 - B. Angiosarcoma
 - C. Palpal purpura (see Table 58-16)
 - D. Primary cutaneous disorders
 - 1. Angiokeratomas of the scrotum and vulva
- X. Brown-black^d
- XI. Any color
 - A. Metastases

^aIf multiple with childhood onset, consider Gardner syndrome. ^bMay have darker hue in more darkly pigmented individuals. ^cMore widespread, especially lower trunk and girdle region, and often red-purple in color. ^dSee also "Hyperpigmentation."

Abbreviations: MEN, multiple endocrine neoplasia; NF1, neurofibromatosis type 1.

periarticular. These patients can also develop calcification of muscular arteries and subsequent ischemic necrosis (calciphylaxis). *Osteoma cutis*, in the form of small papules, most commonly occurs on the face of individuals with a history of acne vulgaris, whereas plate-like lesions occur in rare genetic syndromes.

■ SKIN-COLORED LESIONS

There are several types of skin-colored lesions, including epidermoid cysts, lipomas, rheumatoid nodules, neurofibromas, angiomas, neuromas, and adnexal tumors such as tricholemmomas. Both *epidermoid cysts* and *lipomas* are very common mobile subcutaneous nodules—the former are rubbery and drain cheeselike material (sebum and keratin) if incised. Lipomas are firm and somewhat lobulated on palpation. When extensive facial epidermoid cysts develop during childhood or there is a family history of such lesions, the patient should be examined for other signs of Gardner syndrome, including osteomas and desmoid tumors. *Rheumatoid nodules* are firm 0.5- to 4-cm nodules that favor the extensor aspect of joints, especially the elbows. They are seen in ~20% of patients with rheumatoid arthritis and 6% of patients with Still's disease. Biopsies of the nodules show palisading granulomas. Similar lesions that are smaller and shorter-lived are seen in rheumatic fever.

Neurofibromas (benign Schwann cell tumors) are soft papules or nodules that exhibit the “button-hole” sign; that is, they invaginate into the skin with pressure in a manner similar to a hernia. Single lesions are seen in normal individuals, but multiple neurofibromas, usually in combination with six or more CALMs measuring >1.5 cm (see “Hyperpigmentation,” above), axillary freckling, and multiple Lisch nodules, are seen in von Recklinghausen’s disease (NF type I) (**Chap. 90**). In some patients, the neurofibromas are localized and unilateral due to somatic mosaicism.

Angiofibromas are firm pink-red to skin-colored papules that measure from 3 mm to 1.5 cm in diameter. When multiple lesions are located on the central cheeks (adenoma sebaceum), the patient has tuberous sclerosis or multiple endocrine neoplasia (MEN) syndrome, type 1. The former is an autosomal disorder due to mutations in two different genes, and the associated findings are discussed in the section on ash leaf spots as well as in **Chap. 90**.

Neuromas (benign proliferations of nerve fibers) are also firm, skin-colored papules. They are more commonly found at sites of amputations and in rudimentary polydactyly. However, when there are multiple neuromas on the eyelids, lips, distal tongue, and/or oral mucosa, the patient should be investigated for other signs of MEN syndrome, type 2b. Associated findings include marfanoid habitus, protuberant lips, intestinal ganglioneuromas, and medullary thyroid carcinoma (>75% of patients; **Chap. 388**).

Adnexal tumors are derived from pluripotent cells of the epidermis that can differentiate toward hair, sebaceous, or apocrine or eccrine glands, or remain undifferentiated. *Basal cell carcinomas* (BCCs) are examples of adnexal tumors that have little or no evidence of differentiation. Clinically, they are translucent papules with rolled borders, telangiectasias, and central erosion. BCCs commonly arise in sun-damaged skin of the head and neck as well as the upper trunk. When a patient has multiple BCCs, especially prior to age 30, the possibility of the basal cell nevus syndrome should be raised. It is inherited as an autosomal dominant trait and is associated with jaw cysts, palmar and plantar pits, frontal bossing, medulloblastomas, and calcification of the falx cerebri and diaphragma sellae. *Tricholemmomas* are also skin-colored adnexal tumors but differentiate toward hair follicles and can have a wartlike appearance. The presence of multiple tricholemmomas on the face and cobblestoning of the oral mucosa points to the diagnosis of Cowden disease (multiple hamartoma syndrome) due to mutations in the phosphatase and tensin homolog (*PTEN*) gene. Internal organ involvement (in decreasing order of frequency) includes fibrocystic disease and carcinoma of the breast, adenomas and carcinomas of the thyroid, and gastrointestinal polyposis. Keratoses of the palms, soles, and dorsal aspect of the hands are also seen. *Fibrofolliculomas* are skin-colored to white, smooth papules that favor the face, ears, and neck and, when multiple, are associated

with Birt-Hogg-Dubé syndrome, which is associated with renal lesions including cancer (**Chap. 85**).

■ PINK LESIONS

The cutaneous lesions associated with primary systemic *amyloidosis* are often pink to pink-orange in color and translucent. Common locations are the face, especially the periorbital and perioral regions, and flexural areas. On biopsy, homogeneous deposits of amyloid are seen in the dermis and in the walls of blood vessels; the latter lead to an increase in vessel wall fragility. As a result, petechiae and purpura develop in clinically normal skin as well as in lesional skin following minor trauma, hence the term *pinch purpura*. Amyloid deposits are also seen in the striated muscle of the tongue and result in macroglossia.

Even though specific mucocutaneous lesions are present in only ~30% of the patients with primary systemic (AL) amyloidosis, the diagnosis can be made via histologic examination of abdominal subcutaneous fat, in conjunction with a serum free light chain assay. By special staining, amyloid deposits are seen around blood vessels or individual fat cells in 40–50% of patients. There are also three forms of amyloidosis that are limited to the skin and that should not be construed as cutaneous lesions of systemic amyloidosis. They are macular amyloidosis (upper back), lichen amyloidosis (usually lower extremities), and nodular amyloidosis. In macular and lichen amyloidosis, the deposits are composed of altered epidermal keratin. Early-onset macular and lichen amyloidosis have been associated with MEN syndrome, type 2a.

Patients with *multicentric reticulohistiocytosis* also have pink-colored papules and nodules on the face and mucous membranes as well as on the extensor surface of the hands and forearms. They have a polyarthritides that can mimic rheumatoid arthritis clinically. On histologic examination, the papules have characteristic giant cells that are not seen in biopsies of rheumatoid nodules. Pink to skin-colored papules that are firm, 2–5 mm in diameter, and often in a linear arrangement are seen in patients with *papular mucinosis*. This disease is also referred to as *scleromyxedema*. The latter name comes from the induration of the face and extremities that may accompany the papular eruption. Biopsy specimens of the papules show localized mucin deposition, and serum protein electrophoresis plus immunofixation electrophoresis demonstrates a monoclonal spike of IgG, usually with a λ light chain.

■ YELLOW LESIONS

Several systemic disorders are characterized by yellow-colored cutaneous papules or plaques—hyperlipidemia (xanthomas), gout (tophi), diabetes (necrobiosis lipoidica), pseudoxanthoma elasticum, and Muir-Torre syndrome (sebaceous tumors). Eruptive xanthomas are the most common form of *xanthomas* and are associated with hypertriglyceridemia (primarily hyperlipoproteinemia types I, IV, and V). Crops of yellow papules with erythematous halos occur primarily on the extensor surfaces of the extremities and the buttocks, and they spontaneously involute with a fall in serum triglycerides. Types II and III result in one or more of the following types of xanthoma: xanthelasma, tendon xanthomas, and plane xanthomas. Xanthelasma are found on the eyelids, whereas tendon xanthomas are frequently associated with the Achilles and extensor finger tendons; plane xanthomas are flat and favor the palmar creases and flexural folds. Tuberous xanthomas are frequently associated with hypercholesterolemia; however, they are also seen in patients with hypertriglyceridemia and are found most frequently over the large joints or hand. Biopsy specimens of xanthomas show collections of lipid-containing macrophages (foam cells).

Patients with several disorders, including biliary cirrhosis, can have a secondary form of hyperlipidemia with associated tuberous and plane xanthomas. However, patients with plasma cell dyscrasias have *normolipemic plane xanthomas*. This latter form of xanthoma may be ≥12 cm in diameter and is most frequently seen on the neck, upper trunk, and flexural folds. It is important to note that the most common setting for eruptive xanthomas is uncontrolled diabetes mellitus. The least specific sign for hyperlipidemia is xanthelasma, because at least 50% of the patients with this finding have normal lipid profiles.

In *tophaceous gout*, there are deposits of monosodium urate in the skin around the joints, particularly those of the hands and feet. Additional

sites of *tophi* formation include the helix of the ear and the olecranon and prepatellar bursae. The lesions are firm, yellow to yellow-white in color, and occasionally discharge a chalky material. Their size varies from 1 mm to 7 cm, and the diagnosis can be established by polarized light microscopy of the aspirated contents of a tophus. Lesions of *necrobiosis lipoidica* are found primarily on the shins (90%), and patients can have diabetes mellitus or develop it subsequently. Characteristic findings include a central yellow color, atrophy (transparency), telangiectasias, and a red to red-brown border. Ulcerations can also develop within the plaques. Biopsy specimens show necrobiosis of collagen and granulomatous inflammation.

In *pseudoxanthoma elasticum* (PXE), due to mutations in the gene ABCC6, there is an abnormal deposition of calcium on the elastic fibers of the skin, eye, and blood vessels. In the skin, the flexural areas such as the neck, axillae, antecubital fossae, and inguinal area are the primary sites of involvement. Yellow papules coalesce to form reticulated plaques that have an appearance similar to that of plucked chicken skin. In severely affected skin, hanging, redundant folds develop. Biopsy specimens of involved skin show swollen and irregularly clumped elastic fibers with deposits of calcium. In the eye, the calcium deposits in Bruch's membrane lead to angiod streaks and choroiditis; in the arteries of the heart, kidney, gastrointestinal tract, and extremities, the deposits lead to angina, hypertension, gastrointestinal bleeding, and claudication, respectively.

Adnexal tumors that have differentiated toward sebaceous glands include sebaceous adenoma, sebaceous carcinoma, and sebaceous hyperplasia. Except for sebaceous hyperplasia, which is commonly seen on the face, these tumors are fairly rare. Patients with Muir-Torre syndrome have one or more *sebaceous adenoma(s)*, and they can also have sebaceous carcinomas and sebaceous hyperplasia as well as keratoacanthomas. The internal manifestations of Muir-Torre syndrome include multiple carcinomas of the gastrointestinal tract (primarily colon) as well as cancers of the genitourinary tract.

RED LESIONS

Cutaneous lesions that are red in color have a wide variety of etiologies; in an attempt to simplify their identification, they will be subdivided into papules, papules/plaques, and subcutaneous nodules. Common red papules include *arthropod bites* and *cherry hemangiomas*; the latter are small, bright-red, dome-shaped papules that represent a benign proliferation of capillaries. In patients with AIDS (Chap. 202), the development of multiple red hemangioma-like lesions points to bacillary angiomatosis, and biopsy specimens show clusters of bacilli that stain positively with the Warthin-Starry stain; the pathogens have been identified as *Bartonella henselae* and *Bartonella quintana*. Disseminated visceral disease is seen primarily in immunocompromised hosts but can occur in immunocompetent individuals.

Multiple *angiokeratomas* are seen in Fabry disease, an X-linked recessive lysosomal storage disease that is due to a deficiency of α-galactosidase A. The lesions are red to red-purple in color and can be quite small in size (1–3 mm), with the most common location being the lower trunk. Associated findings include chronic renal disease, peripheral neuropathy, and corneal opacities (cornea verticillata). While electron photomicrographs demonstrate lamellar lipid deposits in dermal fibroblasts, pericytes, and endothelial cells, nowadays, genetic analysis is more frequently performed for diagnosis. Widespread acute eruptions of erythematous papules are discussed in the section on exanthems.

There are several infectious diseases that present as erythematous papules or nodules in a lymphocutaneous or sporotrichoid pattern, that is, in a linear arrangement along the lymphatic channels. The two most common etiologies are *Sporothrix schenckii* (sporotrichosis) and the atypical mycobacterium *Mycobacterium marinum*. The organisms are introduced as a result of trauma, and a primary inoculation site is often seen in addition to the lymphatic nodules. Additional causes include *Nocardia*, *Leishmania*, and other atypical mycobacteria and dimorphic fungi; culture or PCR of lesional tissue will aid in the diagnosis.

The diseases that are characterized by erythematous plaques with scale are reviewed in the papulosquamous section, and the various

forms of dermatitis are discussed in the section on erythroderma. Additional disorders in the differential diagnosis of red papules/plaques include *cellulitis*, *polymorphous light eruption* (PMLE), *cutaneous lymphoid hyperplasia* (lymphocytoma cutis), *cutaneous lupus*, *lymphoma cutis*, and *leukemia cutis*. The first three diseases represent primary cutaneous disorders, although cellulitis may be accompanied by a bacteremia. PMLE is characterized by erythematous papules and plaques in a primarily sun-exposed distribution—dorsum of the hand, extensor forearm, and upper trunk. Lesions follow exposure to UV-B and/or UV-A, and in higher latitudes, PMLE is most severe in the late spring and early summer. A process referred to as “hardening” occurs with continued UV exposure, and the eruption fades, but in temperate climates, it recurs the next spring. PMLE must be differentiated from cutaneous lupus, and this is accomplished by observation of the natural history, histologic examination, and sometimes direct immunofluorescence of the lesions. Cutaneous lymphoid hyperplasia (pseudolymphoma) is a *benign* polyclonal proliferation of lymphocytes within the skin that presents as infiltrated pink-red to red-purple papules and plaques; it must be distinguished from lymphoma cutis.

Several types of red plaques are seen in patients with systemic *lupus*, including (1) erythematous urticarial plaques across the cheeks and nose in the classic butterfly rash; (2) erythematous discoid lesions with fine or “carpet-tack” scale, telangiectasias, central hypopigmentation, peripheral hyperpigmentation, follicular plugging, and atrophy located on the scalp, face, external ears, arms, and upper trunk; and (3) psoriasiform or annular lesions of subacute cutaneous lupus with hypopigmented centers located primarily on the extensor arms and upper trunk. Additional mucocutaneous findings include (1) a violaceous flush on the face and V of the neck; (2) photosensitivity; (3) urticarial vasculitis (see “Urticaria,” above); (4) lupus panniculitis (see below); (5) diffuse alopecia; (6) alopecia secondary to discoid lesions; (7) nailfold telangiectasias and erythema; (8) EM- or TEN-like lesions that may become bullous; (9) oral or nasal ulcers; (10) livedo reticularis; and (11) distal ulcerations secondary to Raynaud’s phenomenon, vasculitis, or livedoid vasculopathy. Patients with only discoid lesions usually have the form of lupus that is limited to the skin. However, up to 10–15% of these patients eventually develop systemic lupus. Direct immunofluorescence of involved skin, in particular discoid lesions, shows deposits of IgG or IgM and C3 in a granular distribution along the dermal-epidermal junction.

In *lymphoma cutis*, there is a clonal proliferation of malignant lymphocytes within the skin, and the clinical appearance resembles that of cutaneous lymphoid hyperplasia—infiltrated pink-red to red-purple papules and plaques. Lymphoma cutis can occur anywhere on the surface of the skin, whereas the sites of predilection for lymphocytomas include the malar ridge, tip of the nose, and earlobes. Patients with non-Hodgkin’s lymphomas have specific cutaneous lesions more often than those with Hodgkin’s lymphoma, and, occasionally, the skin nodules precede the development of extracutaneous non-Hodgkin’s lymphoma or represent the only site of involvement (e.g., primary cutaneous B-cell lymphoma). Arcuate lesions are sometimes seen in lymphoma and lymphocytoma cutis as well as in CTCL. Adult *T-cell leukemia/lymphoma* that develops in association with HTLV-1 infection is characterized by cutaneous plaques, hypercalcemia, and circulating CD25+ lymphocytes. *Leukemia cutis* has the same appearance as lymphoma cutis, and specific lesions are seen more commonly in monocytic leukemias than in lymphocytic or granulocytic leukemias. Cutaneous chloromas (granulocytic sarcomas) may precede the appearance of circulating blasts in acute myelogenous leukemia and, as such, represent a form of aleukemic leukemia cutis.

Sweet syndrome is characterized by pink-red to red-brown edematous plaques that are frequently painful and occur primarily on the head, neck, and upper extremities. The patients also have fever, neutrophilia, and a dense dermal infiltrate of neutrophils in the lesions. In ~10% of the patients, there is an associated malignancy, most commonly acute myelogenous leukemia. Sweet syndrome has also been reported with inflammatory bowel disease, systemic lupus erythematosus, and solid tumors (primarily of the genitourinary tract) as well as drugs (e.g., granulocyte colony-stimulating factor [G-CSF], hypomethylating

agents, all-*trans*-retinoic acid). The differential diagnosis includes neutrophilic eccrine hidradenitis; bullous forms of pyoderma gangrenosum; and, occasionally, cellulitis. Extracutaneous sites of involvement include joints, muscles, eyes, kidneys (proteinuria, occasionally glomerulonephritis), and lungs (neutrophilic infiltrates). The idiopathic form of Sweet syndrome is seen more often in women, following a respiratory tract infection.

Common causes of erythematous subcutaneous nodules include inflamed epidermoid cysts, acne cysts, and furuncles. *Panniculitis*, an inflammation of the fat, also presents as subcutaneous nodules and is frequently a sign of systemic disease. There are several forms of panniculitis, including erythema nodosum, erythema induratum/nodular vasculitis, lupus panniculitis, lipodermatosclerosis, α_1 -antitrypsin deficiency, factitial, and fat necrosis secondary to pancreatic disease. Except for erythema nodosum, these lesions may break down and ulcerate or heal with a scar. The shin is the most common location for the nodules of erythema nodosum, whereas the calf is the most common location for lesions of erythema induratum. In erythema nodosum, the nodules are initially red but then develop a blue bruise-like color as they resolve. Patients with erythema nodosum but no underlying systemic illness can still have fever, malaise, leukocytosis, arthralgias, and/or arthritis. However, the possibility of an underlying illness should be excluded, and the most common associations are streptococcal infections, upper respiratory viral infections, sarcoidosis, and inflammatory bowel disease, in addition to drugs (oral contraceptives, sulfonamides, penicillins, bro-mides, iodides, BRAF inhibitors). Less common associations include bacterial gastroenteritis (*Yersinia*, *Salmonella*) and coccidioidomycosis followed by tuberculosis, histoplasmosis, brucellosis, and infections with *Chlamydia pneumoniae*, *Chlamydia trachomatis*, *Mycoplasma pneumoniae*, or hepatitis B virus.

Erythema induratum and nodular vasculitis have overlapping features clinically and histologically, and whether they represent two separate entities or the ends of a single disease spectrum is a point of debate; in general, the latter is usually idiopathic and the former is associated with the presence of *Mycobacterium tuberculosis* DNA by PCR within skin lesions. The lesions of lupus panniculitis are found primarily on the cheeks, upper arms, and buttocks (sites of abundant fat) and are seen in both the cutaneous and systemic forms of lupus. The overlying skin may be normal, erythematous, or have the changes of discoid lupus. The subcutaneous fat necrosis that is associated with pancreatic disease is presumably secondary to circulating lipases and is seen in patients with pancreatic carcinoma as well as in patients with acute and chronic pancreatitis. In this disorder, there may be an associated arthritis, fever, and inflammation of visceral fat. Histologic examination of deep incisional biopsy specimens will aid in the diagnosis of the particular type of panniculitis.

Subcutaneous erythematous nodules are also seen in cutaneous polyarteritis nodosa and as a manifestation of *systemic vasculitis* when there is involvement of medium-sized vessels, for example, systemic polyarteritis nodosa, eosinophilic granulomatosis with polyangiitis, or granulomatosis with polyangiitis (Chap. 363). Cutaneous polyarteritis nodosa presents with painful subcutaneous nodules and ulcers within a red-purple, netlike pattern of livedo reticularis. The latter is due to slowed blood flow through the superficial horizontal venous plexus. The majority of lesions are found on the lower extremities, and while arthralgias and myalgias may accompany cutaneous polyarteritis nodosa, there is no evidence of systemic involvement. In both the cutaneous and systemic forms of vasculitis, skin biopsy specimens of the associated nodules will show the changes characteristic of a necrotizing vasculitis and/or granulomatous inflammation.

■ RED-BROWN LESIONS

The cutaneous lesions in *sarcoidosis* (Chap. 367) are classically red to red-brown in color, and with diascopy (pressure with a glass slide), a yellow-brown residual color is observed that is secondary to the granulomatous infiltrate. The waxy papules and plaques may be found anywhere on the skin, but the face is the most common location. Usually there are no surface changes, but occasionally, the lesions will have scale. Biopsy specimens of the papules show "naked" granulomas in

the dermis, that is, granulomas surrounded by a minimal number of lymphocytes. Other cutaneous findings in sarcoidosis include annular lesions with an atrophic or scaly center, papules within scars, hypopigmented papules and patches, subcutaneous plaques, alopecia, acquired ichthyosis, erythema nodosum, and lupus pernio (see below).

The differential diagnosis of sarcoidosis includes foreign-body granulomas produced by chemicals such as beryllium and zirconium, late secondary syphilis, and *lupus vulgaris*. Lupus vulgaris is a form of cutaneous tuberculosis that is seen in previously infected and sensitized individuals. There is often underlying active tuberculosis elsewhere, usually in the lungs or lymph nodes. Lesions occur primarily in the head and neck region and are red-brown plaques with a yellow-brown color on diascopy. Secondary scarring can develop within the central portion of the plaques. Cultures or PCR analysis of the lesions should be performed, along with an interferon γ release assay of peripheral blood, because it is rare for the acid-fast stain to show bacilli within the dermal granulomas.

A generalized distribution of red-brown macules and papules is seen in the form of mastocytosis known as *urticaria pigmentosa* (Chap. 354). Each lesion represents a collection of mast cells in the dermis, with hyperpigmentation of the overlying epidermis. Stimuli such as rubbing cause these mast cells to degranulate, and this leads to the formation of localized urticaria (Darier's sign). Additional symptoms can result from mast cell degranulation and include headache, flushing, diarrhea, and pruritus. Mast cells also infiltrate various organs such as the liver, spleen, and gastrointestinal tract, and accumulations of mast cells in the bones may produce either osteosclerotic or osteolytic lesions on radiographs. In the majority of these patients, however, the internal involvement remains indolent. A subtype of chronic cutaneous small-vessel vasculitis, *erythema elevatum diutinum* (EED), also presents with papules that are red-brown in color. The papules coalesce into plaques on the extensor surfaces of knees, elbows, and the small joints of the hand. Flares of EED have been associated with streptococcal infections.

■ BLUE LESIONS

Lesions that are blue in color are the result of vascular ectasias, hyperplasias, and tumors or melanin pigment within the dermis. *Venous lakes* (ectasias) are compressible dark-blue lesions that are found commonly in the head and neck region. *Venous malformations* are also compressible blue papulonodules and plaques that can occur anywhere on the body, including the oral mucosa. When there are multiple papulonodules rather than a single congenital lesion, the patient may have the blue rubber bleb syndrome or Maffucci's syndrome. Patients with the blue rubber bleb syndrome also have vascular anomalies of the gastrointestinal tract that may bleed, whereas patients with Maffucci's syndrome have associated osteochondromas. *Blue nevi* (moles) are seen when there are collections of pigment-producing nevus cells in the dermis. These benign papular lesions are dome-shaped and occur most commonly on the dorsum of the hand or foot or in the head and neck region.

■ VIOLACEOUS LESIONS

Violaceous papules and plaques are seen in *lupus pernio*, *lymphoma cutis*, and *cutaneous lupus*. Lupus pernio is a particular type of sarcoidosis that involves the tip and alar rim of the nose as well as the earlobes, with lesions that are violaceous in color rather than red-brown. This form of sarcoidosis is associated with involvement of the upper respiratory tract. The plaques of lymphoma cutis and cutaneous lupus may be red or violaceous in color and were discussed above.

■ PURPLE LESIONS

Purple-colored papules and plaques are seen in vascular tumors, such as *Kaposi's sarcoma* (Chap. 202) and *angiosarcoma*, and when there is extravasation of red blood cells into the skin in association with inflammation, as in *palpable purpura* (see "Purpura," below). Patients with congenital or acquired AV fistulas and venous hypertension can develop purple papules on the lower extremities that can resemble Kaposi's sarcoma clinically and histologically; this condition is referred

to as pseudo-Kaposi's sarcoma (acral angiodermatitis). Angiosarcoma is found most commonly on the scalp and face of elderly patients or within areas of chronic lymphedema and presents as purple papules and plaques. In the head and neck region, the tumor often extends beyond the clinically defined borders and may be accompanied by facial edema.

BROWN AND BLACK LESIONS

Brown- and black-colored papules are reviewed in "Hyperpigmentation," above.

CUTANEOUS METASTASES

These are discussed last because they can have a wide range of colors. Most commonly, they present as either firm, skin-colored subcutaneous nodules or firm, red to red-brown papulonodules, whereas metastatic melanoma can be pink, blue, or black in color. Cutaneous metastases develop from hematogenous or lymphatic spread and are most often due to the following primary carcinomas: in men, melanoma, oropharynx, lung, and colon; and in women, breast, melanoma, and ovary. These metastatic lesions may be the initial presentation of the carcinoma, especially when the primary site is the lung.

PURPURA

(Table 58-16) *Purpura* are seen when there is an extravasation of red blood cells into the dermis and, as a result, the lesions do not blanch with pressure. This is in contrast to those erythematous or violet-colored lesions that are due to localized vasodilatation—they do blanch with pressure. Purpura (≥ 3 mm) and petechiae (≤ 2 mm) are divided into two major groups: palpable and nonpalpable. The most frequent causes of *nonpalpable* purpura and petechiae are primary cutaneous disorders such as *trauma*, *solar (actinic) purpura*, *stasis purpura*, and *capillaritis*. Less common causes are *steroid purpura* and *livedoid vasculopathy* (see "Ulcers," below). Solar purpura are seen primarily on the extensor forearms, whereas steroid purpura secondary to potent topical glucocorticoids or endogenous or exogenous Cushing's syndrome can be more widespread. In both cases, there is alteration of the supporting connective tissue that surrounds the dermal blood vessels. In contrast, the petechiae that result from capillaritis are found primarily on the lower extremities. In capillaritis, there is an extravasation of erythrocytes as a result of perivascular lymphocytic inflammation. The petechiae are bright red, 1–2 mm in size, and scattered within yellow-brown patches. The yellow-brown color is caused by hemosiderin deposits within the dermis.

Systemic causes of nonpalpable purpura fall into several categories, and those secondary to clotting disturbances and vascular fragility will be discussed first. The former group includes *thrombocytopenia* (Chap. 115), *abnormal platelet function* as is seen in uremia, and *clotting factor defects*. The initial site of presentation for thrombocytopenia-induced petechiae is the distal lower extremity. Capillary fragility leads to nonpalpable purpura in patients with systemic *amyloidosis* (see "Papulonodular Skin Lesions," above), disorders of collagen production such as *Ehlers-Danlos syndrome*, and *scurvy*. In scurvy, there are flattened corkscrew hairs with surrounding hemorrhage on the lower extremities, in addition to gingivitis. Vitamin C is a cofactor for lysyl hydroxylase, an enzyme involved in the posttranslational modification of procollagen that is necessary for cross-link formation.

In contrast to the previous group of disorders, the noninflammatory purpura seen in the following group of diseases are associated with thrombi formation within vessels and have a retiform configuration. It is important to note that these thrombi are demonstrable in skin biopsy specimens. This group of disorders includes disseminated intravascular coagulation (DIC), monoclonal cryoglobulinemia, thrombocytosis, thrombotic thrombocytopenic purpura, antiphospholipid antibody syndrome, and reactions to warfarin and heparin (heparin-induced thrombocytopenia and thrombosis). DIC is triggered by several types of infection (gram-negative, gram-positive, viral, and rickettsial) as well as by tissue injury and neoplasms. Widespread purpura and hemorrhagic infarcts of the distal extremities are seen. Similar lesions are found in purpura fulminans, which is a form of DIC associated with

TABLE 58-16 Causes of Purpura

- I. Primary cutaneous disorders
 - A. Nonpalpable
 - 1. Trauma
 - 2. Solar (actinic, senile) purpura
 - 3. Steroid purpura
 - 4. Stasis purpura due to venous hypertension
 - 5. Capillaritis
 - 6. Livedoid vasculopathy in the setting of venous hypertension^a
 - B. Drugs (e.g., antiplatelet agents, anticoagulants)
 - C. Systemic diseases
 - A. Nonpalpable
 - 1. Clotting disturbances
 - a. Thrombocytopenia (including ITP)
 - b. Abnormal platelet function
 - c. Clotting factor defects
 - 2. Vascular fragility
 - a. Amyloidosis (within normal-appearing skin)
 - b. Ehlers-Danlos syndrome
 - c. Scurvy
 - 3. Thrombi
 - a. Disseminated intravascular coagulation, purpura fulminans
 - b. Warfarin (Coumadin)-induced necrosis
 - c. Heparin-induced thrombocytopenia and thrombosis
 - d. Antiphospholipid antibody syndrome
 - e. Monoclonal cryoglobulinemia
 - f. Vasculopathy induced by levamisole-adulterated cocaine^b
 - g. SARS-CoV-2 infection
 - h. Thrombotic thrombocytopenic purpura
 - i. Thrombocytosis
 - j. Homozygous protein C or protein S deficiency
 - 4. Emboli
 - a. Cholesterol
 - b. Fat
 - 5. Possible immune complex
 - a. Gardner-Diamond syndrome (autoerythrocyte sensitivity)
 - b. Waldenström's hypergammaglobulinemic purpura
 - B. Palpable
 - 1. Vasculitis
 - a. Cutaneous small-vessel vasculitis, including in the setting of systemic vasculitides
 - 2. Emboli^c
 - a. Acute meningococcemia
 - b. Disseminated gonococcal infection
 - c. Rocky Mountain spotted fever
 - d. Ecthyma gangrenosum

^aAlso associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency. ^bCombined vasculopathy/vasculitis can be seen. ^cBacterial (including rickettsial), fungal, or parasitic.

Abbreviation: ITP, idiopathic thrombocytopenic purpura.

fever and hypotension that occurs more commonly in children following an infectious illness such as varicella, scarlet fever, or an upper respiratory tract infection. In both disorders, hemorrhagic bullae can develop in involved skin.

Monoclonal cryoglobulinemia is associated with plasma cell dyscrasias, chronic lymphocytic leukemia, and lymphoma. Purpura, primarily of the lower extremities, and hemorrhagic infarcts of the fingers, toes, nose and ears are seen in these patients. Exacerbations of disease activity can follow cold exposure or an increase in serum viscosity. Biopsy specimens show precipitates of the cryoglobulin within dermal vessels. Similar deposits have been found in the lung, brain, and renal glomeruli. Patients with *thrombotic thrombocytopenic purpura* can also have hemorrhagic infarcts as a result of intravascular thromboses.

Additional signs include microangiopathic hemolytic anemia and fluctuating neurologic abnormalities, especially headaches and confusion.

Administration of warfarin can result in painful areas of erythema that become purpuric and then necrotic with an adherent black eschar; the condition is also referred to as Coumadin-induced necrosis. This reaction is seen more often in women and in areas with abundant subcutaneous fat—breasts, abdomen, buttocks, thighs, and calves. The erythema and purpura develop between the third and tenth day of therapy, most likely as a result of a transient imbalance in the levels of anticoagulant and procoagulant vitamin K-dependent factors. Continued therapy does not exacerbate preexisting lesions, and patients with an inherited or acquired deficiency of protein C are at increased risk for this particular reaction as well as for purpura fulminans and calciphylaxis.

Purpura secondary to *cholesterol emboli* are usually seen on the lower extremities of patients with atherosclerotic vascular disease. They often follow anticoagulant therapy or an invasive vascular procedure such as an arteriogram but also occur spontaneously from disintegration of atheromatous plaques. Associated findings include livedo reticularis, gangrene, cyanosis, and ischemic ulcerations. Multiple step sections of the biopsy specimen may be necessary to demonstrate the cholesterol clefts within the vessels. Petechiae are also an important sign of *fat embolism* and occur primarily on the upper body 2–3 days after a major injury. By using special fixatives, the emboli can be demonstrated in biopsy specimens of the petechiae. Rarely, emboli of tumor or thrombus are seen in patients with atrial myxomas and marantic endocarditis.

In the *Gardner-Diamond syndrome* (autoerythrocyte sensitivity), female patients develop large ecchymoses within areas of painful, warm erythema. Intradermal injections of autologous erythrocytes or phosphatidyl serine derived from the red cell membrane can reproduce the lesions in some patients; however, there are instances where a reaction is seen at an injection site of the forearm but not in the midback region. The latter has led some observers to view Gardner-Diamond syndrome as a cutaneous manifestation of severe emotional stress. More recently, the possibility of platelet dysfunction (as assessed via aggregation studies) has been raised. *Waldenström's hypergammaglobulinemic purpura* is a chronic disorder characterized by recurrent crops of petechiae and larger purpuric macules on the lower extremities. There are circulating complexes of IgG-anti-IgG molecules, and exacerbations are associated with prolonged standing or walking. Patients may have an underlying autoimmune connective tissue disease, e.g., Sjögren's syndrome.

Palpable purpura are further subdivided into vasculitic and embolic. In the group of vasculitic disorders, cutaneous small-vessel vasculitis, also known as *leukocytoclastic vasculitis* (LCV), is the one most commonly associated with palpable purpura (Chap. 363). Underlying etiologies include drugs (e.g., antibiotics), infections (e.g., hepatitis C virus), and autoimmune connective tissue diseases (e.g., rheumatoid arthritis, Sjögren's syndrome, lupus). *Henoch-Schönlein purpura* (HSP) is a subtype of acute LCV that is seen more commonly in children and adolescents following an upper respiratory infection. The majority of lesions are found on the lower extremities and buttocks. Systemic manifestations include fever, arthralgias (primarily of the knees and ankles), abdominal pain, gastrointestinal bleeding, and nephritis. Direct immunofluorescence examination shows deposits of IgA within dermal blood vessel walls. Renal disease is of particular concern in adults with IgA vasculitis.

Several types of infectious emboli can give rise to palpable purpura. These embolic lesions are usually *irregular* in outline as opposed to the lesions of LCV, which are *circular* in outline. The irregular outline is indicative of a cutaneous infarct, and the size corresponds to the area of skin that received its blood supply from that particular arteriole or artery. The palpable purpura in LCV are circular because the erythrocytes simply diffuse out evenly from the postcapillary venules as a result of inflammation. Infectious emboli are most commonly due to gram-negative cocci (meningococcus, gonococcus), gram-negative rods (Enterobacteriaceae), and gram-positive cocci (*Staphylococcus*). Additional causes include *Rickettsia* and, in immunocompromised patients, *Aspergillus* and other opportunistic fungi.

The embolic lesions in *acute meningococcemia* are found primarily on the trunk, lower extremities, and sites of pressure, and a gunmetal-gray color often develops within them. Their size varies from a few millimeters to several centimeters, and the organisms can be cultured from the lesions. Associated findings include a preceding upper respiratory tract infection; fever; meningitis; DIC; and, in some patients, a deficiency of the terminal components of complement. In *disseminated gonococcal infection* (arthritis-dermatitis syndrome), a small number of inflammatory papules and vesicopustules, often with central purpura or hemorrhagic necrosis, are found on the distal extremities. Additional symptoms include arthralgias, tenosynovitis, and fever. To establish the diagnosis, a Gram stain of these lesions should be performed. *Rocky Mountain spotted fever* is a tick-borne disease that is caused by *Rickettsia rickettsii*. A several-day history of fever, chills, severe headache, and photophobia precedes the onset of the cutaneous eruption. The initial lesions are erythematous macules and papules on the wrists, ankles, palms, and soles. With time, the lesions spread centripetally and become purpuric.

Lesions of *ecthyma gangrenosum* begin as edematous, erythematous papules or plaques and then develop central purpura and necrosis. Bullae formation also occurs in these lesions, and they are frequently found in the girdle region. The organism that is classically associated with ecthyma gangrenosum is *Pseudomonas aeruginosa*, but other gram-negative rods such as *Klebsiella*, *Escherichia coli*, and *Serratia* can produce similar lesions. In immunocompromised hosts, the list of potential pathogens is expanded to include *Candida* and other opportunistic fungi (e.g., *Aspergillus*, *Fusarium*).

ULCERS

The approach to the patient with a cutaneous ulcer is outlined in Table 58-17. Peripheral vascular diseases of the extremities are reviewed in Chap. 281, as is Raynaud's phenomenon.

Livedoid vasculopathy (livedoid vasculitis; atrophie blanche) represents a combination of a vasculopathy plus intravascular thrombosis. Purpuric lesions and livedo reticularis are found in association with *painful* ulcerations of the lower extremities. These ulcers are often slow to heal, but when they do, irregularly shaped white scars form. The majority of cases are secondary to venous hypertension, but possible underlying illnesses include disorders of hypercoagulability, for example, antiphospholipid syndrome and factor V Leiden (Chaps. 117 and 357).

In *pyoderma gangrenosum*, the border of untreated active ulcers has a characteristic appearance consisting of an undermined necrotic violaceous edge and a peripheral erythematous halo. The ulcers often begin as pustules that then expand rather rapidly to a size as large as 20 cm. Although these lesions are most commonly found on the lower extremities, they can arise anywhere on the surface of the body, including at sites of trauma (pathergy). An estimated 30–50% of cases are idiopathic, and the most common associated disorders are ulcerative colitis and Crohn's disease. Less commonly, pyoderma gangrenosum is associated with seropositive rheumatoid arthritis, acute and chronic myelogenous leukemia, myelodysplasia, a monoclonal gammopathy (usually IgA), or an autoinflammatory disorder. Because the histology of pyoderma gangrenosum may be nonspecific (dermal infiltrate of neutrophils when in untreated state), the diagnosis requires clinicopathologic correlation, in particular, the exclusion of similar-appearing ulcers such as necrotizing vasculitis, Meleney's ulcer (synergistic infection at a site of trauma or surgery), dimorphic fungi, cutaneous amebiasis, spider bites, and factitial. In the myeloproliferative disorders, the ulcers may be more superficial with a pustulobullous border, and these lesions provide a connection between classic pyoderma gangrenosum and acute febrile neutrophilic dermatosis (Sweet syndrome).

FEVER AND RASH

The major considerations in a patient with a fever and a rash are inflammatory diseases versus infectious diseases. In the hospital setting, the most common scenario is a patient who has a drug rash plus a fever secondary to an underlying infection. However, it should be emphasized that a drug reaction can lead to both a cutaneous eruption and a fever ("drug fever"), especially in the setting of DRESS, AGEP, or serum

TABLE 58-17 Causes of Mucocutaneous Ulcers

- I. Primary cutaneous disorders
 - A. Peripheral vascular disease (**Chap. 281**)
 - 1. Venous
 - 2. Arterial^a
 - B. Livedoid vasculopathy in the setting of venous hypertension^b
 - C. Squamous cell carcinoma (e.g., within scars), basal cell carcinomas
 - D. Infections, e.g., ecthyma caused by *Streptococcus* (**Chap. 148**)
 - E. Physical, e.g., trauma, pressure
 - F. Drugs, e.g., hydroxyurea
- II. Systemic diseases
 - A. Lower legs
 - 1. Small-vessel and medium-vessel vasculitis^c
 - 2. Hemoglobinopathies (**Chap. 98**)
 - 3. Cryoglobulinemia,^c cryofibrinogenemia
 - 4. Cholesterol emboli^{a,c}
 - 5. Necrobiosis lipoidica^d
 - 6. Antiphospholipid syndrome (**Chap. 116**)
 - 7. Neuropathic^e (**Chap. 403**)
 - 8. Panniculitis
 - 9. Kaposi's sarcoma, acral angiodermatitis (pseudo-Kaposi's sarcoma)
 - 10. Diffuse dermal angiogenesis
 - B. Hands and feet
 - 1. Raynaud's phenomenon (**Chap. 281**)
 - 2. Buerger disease
 - C. Generalized
 - 1. Pyoderma gangrenosum, but most commonly legs
 - 2. Calciphylaxis (**Chap. 410**)
 - 3. Infections, e.g., dimorphic fungi, leishmaniasis
 - 4. Lymphoma
 - D. Face, especially perioral, and anogenital
 - 1. Chronic herpes simplex^f
- III. Mucosal
 - A. Aphthae
 - B. Drug-induced mucositis
 - C. Behcet's disease (**Chap. 364**)
 - D. Erythema multiforme major, Stevens-Johnson syndrome, TEN
 - E. Primary blistering disorders (**Chap. 59**)
 - F. Lupus erythematosus, lichen planus, lichenoid GVHD
 - G. Inflammatory bowel disease
 - H. Acute HIV infection
 - I. Reactive arthritis

^aUnderlying atherosclerosis. ^bAlso associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency, antiphospholipid antibodies. ^cReviewed in section on purpura. ^dReviewed in section on papulonodular skin lesions. ^eFavors plantar surface of the foot. ^fSign of immunosuppression.

Abbreviations: GVHD, graft versus host disease; HIV, human immunodeficiency virus; TEN, toxic epidermal necrolysis.

sickness-like reaction. Additional inflammatory diseases that are often associated with a fever include pustular psoriasis, erythroderma, and Sweet syndrome. Lyme disease, secondary syphilis, and viral and bacterial exanthems (see “Exanthems,” above) are examples of infectious diseases that produce a rash and a fever. Lastly, it is important to determine whether or not the cutaneous lesions represent septic emboli (see “Purpura,” above). Such lesions usually have evidence of ischemia in the form of purpura, necrosis, or impending necrosis (gunmetal-gray color). In the patient with thrombocytopenia, however, purpura can be seen in inflammatory reactions such as morbilliform drug eruptions and infectious lesions.

FURTHER READING

BOLOGNA JL, SCHAFER JV, CERRONI L (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.

CALLEN JP et al (eds): *Dermatological Signs of Systemic Disease*, 5th ed. Edinburgh, Elsevier, 2017.

FAZEL N (ed): *Oral Signs of Systemic Disease*. Switzerland, Springer, 2019.

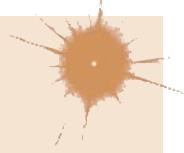
KURTZMAN D: Rheumatologic dermatology. *Clin Dermatol* 36:439, 2018.

TAYLOR SC et al (eds): *Taylor and Kelly's Dermatology for Skin of Color*, 2nd ed. New York, McGraw-Hill, 2016.

59

Immunologically Mediated Skin Diseases

Kim B. Yancey, Benjamin F. Chong,
Thomas J. Lawley



A number of immunologically mediated skin diseases and immunologically mediated systemic disorders with cutaneous manifestations are now recognized as distinct entities with consistent clinical, histologic, and immunopathologic findings. Clinically, these disorders are characterized by morbidity (pain, pruritus, disfigurement) and, in some instances, result in death (largely due to loss of epidermal barrier function and/or secondary infection). The major features of the more common immunologically mediated skin diseases are summarized in this chapter (**Table 59-1**), as are autoimmune systemic disorders with cutaneous manifestations.

AUTOIMMUNE CUTANEOUS DISEASES

■ PEMPHIGUS VULGARIS

Pemphigus refers to a group of autoantibody-mediated intraepidermal blistering diseases characterized by loss of cohesion between epidermal cells (a process termed *acantholysis*). Manual pressure to the skin of these patients may elicit the separation of the epidermis (*Nikolsky's sign*). This finding, while characteristic of pemphigus, is not specific to this group of disorders and is also seen in toxic epidermal necrolysis, Stevens-Johnson syndrome, and a few other skin diseases.

Pemphigus vulgaris (PV) is a mucocutaneous blistering disease that predominantly occurs in patients >40 years of age. PV typically begins on mucosal surfaces and often progresses to involve the skin. This disease is characterized by fragile, flaccid blisters that rupture to produce extensive denudation of mucous membranes and skin (**Fig. 59-1**). The mouth, scalp, face, neck, axilla, groin, and trunk are typically involved. PV may be associated with severe skin pain; some patients experience pruritus as well. Lesions usually heal without scarring except at sites complicated by secondary infection or mechanically induced dermal wounds. Postinflammatory hyperpigmentation is usually present for some time at sites of healed lesions.

Biopsies of early lesions demonstrate intraepidermal vesicle formation secondary to loss of cohesion between epidermal cells (i.e., acantholytic blisters). Blister cavities contain acantholytic epidermal cells, which appear as round homogeneous cells containing hyperchromatic nuclei. Basal keratinocytes remain attached to the epidermal basement membrane; hence, blister formation takes place within the suprabasal portion of the epidermis. Lesional skin may contain focal collections of intraepidermal eosinophils within blister cavities; dermal alterations are slight, often limited to an eosinophil-predominant leukocytic infiltrate. Direct immunofluorescence microscopy of lesional or intact patient skin shows deposits of IgG on the surface of keratinocytes; deposits of complement components are typically found in lesional but not in uninvolved skin. Deposits of IgG on keratinocytes are derived from circulating autoantibodies to cell-surface autoantigens.

TABLE 59-1 Immunologically Mediated Blistering Diseases

DISEASE	CLINICAL MANIFESTATIONS	HISTOLOGY	IMMUNOPATHOLOGY	AUTOANTIGENS ^a
Pemphigus vulgaris	Flaccid blisters, denuded skin, oromucosal lesions	Acantholytic blister formed in suprabasal layer of epidermis	Cell surface deposits of IgG on keratinocytes	Dsg3 (plus Dsg1 in patients with skin involvement)
Pemphigus foliaceus	Crusts and shallow erosions on scalp, central face, upper chest, and back	Acantholytic blister formed in superficial layer of epidermis	Cell surface deposits of IgG on keratinocytes	Dsg1
Paraneoplastic pemphigus	Painful stomatitis with papulosquamous or lichenoid eruptions that may progress to blisters	Acantholysis, keratinocyte necrosis, and vacuolar interface dermatitis	Cell surface deposits of IgG and C3 on keratinocytes and (variably) similar immunoreactants in epidermal BMZ	Plakin protein family members and desmosomal cadherins (see text for details)
Bullous pemphigoid	Large tense blisters on flexor surfaces and trunk	Subepidermal blister with eosinophil-rich infiltrate	Linear band of IgG and/or C3 in epidermal BMZ	BPAG1, BPAG2
Pemphigoid gestationis	Pruritic, urticarial plaques rimmed by vesicles and bullae on the trunk and extremities	Teardrop-shaped, subepidermal blisters in dermal papillae; eosinophil-rich infiltrate	Linear band of C3 in epidermal BMZ	BPAG2 (plus BPAG1 in some patients)
Dermatitis herpetiformis	Extremely pruritic small papules and vesicles on elbows, knees, buttocks, and posterior neck	Subepidermal blister with neutrophils in dermal papillae	Granular deposits of IgA in dermal papillae	Epidermal transglutaminase
Linear IgA disease	Pruritic small papules on extensor surfaces; occasionally larger, arciform blisters	Subepidermal blister with neutrophil-rich infiltrate	Linear band of IgA in epidermal BMZ	BPAG2 (see text for specific details)
Epidermolysis bullosa acquisita	Blisters, erosions, scars, and milia on sites exposed to trauma; widespread, inflammatory, tense blisters may be seen initially	Subepidermal blister that may or may not include a leukocytic infiltrate	Linear band of IgG and/or C3 in epidermal BMZ	Type VII collagen
Mucous membrane pemphigoid	Erosive and/or blistering lesions of mucous membranes and possibly the skin; scarring of some sites	Subepidermal blister that may or may not include a leukocytic infiltrate	Linear band of IgG, IgA, and/or C3 in epidermal BMZ	BPAG2, laminin-332, or others

^aAutoantigens bound by these patients' autoantibodies are defined as follows: Dsg1, desmoglein 1; Dsg3, desmoglein 3; BPAG1, bullous pemphigoid antigen 1; BPAG2, bullous pemphigoid antigen 2.

Abbreviation: BMZ, basement membrane zone.

Such circulating autoantibodies can be demonstrated in 80–90% of PV patients by indirect immunofluorescence microscopy; monkey esophagus is the optimal substrate for these studies. Patients with PV have IgG autoantibodies to *desmogleins* (Dsgs), transmembrane desmosomal glycoproteins that belong to the cadherin family of calcium-dependent adhesion molecules. Such autoantibodies can be precisely quantitated by enzyme-linked immunosorbent assay (ELISA). Patients with early PV (i.e., mucosal disease) have IgG autoantibodies to Dsg3; patients with advanced PV (i.e., mucocutaneous disease) have

IgG autoantibodies to both Dsg3 and Dsg1. Experimental studies have shown that autoantibodies from patients with PV are pathogenic (i.e., responsible for blister formation) and that their titer correlates with disease activity. Recent studies have shown that the anti-Dsg antibody profile in these patients' sera as well as the tissue distribution of Dsg3 and Dsg1 determine the site of blister formation in patients with PV. Coexpression of Dsg3 and Dsg1 by epidermal cells protects against pathogenic IgG antibodies to either of these cadherins but not against pathogenic autoantibodies to both.

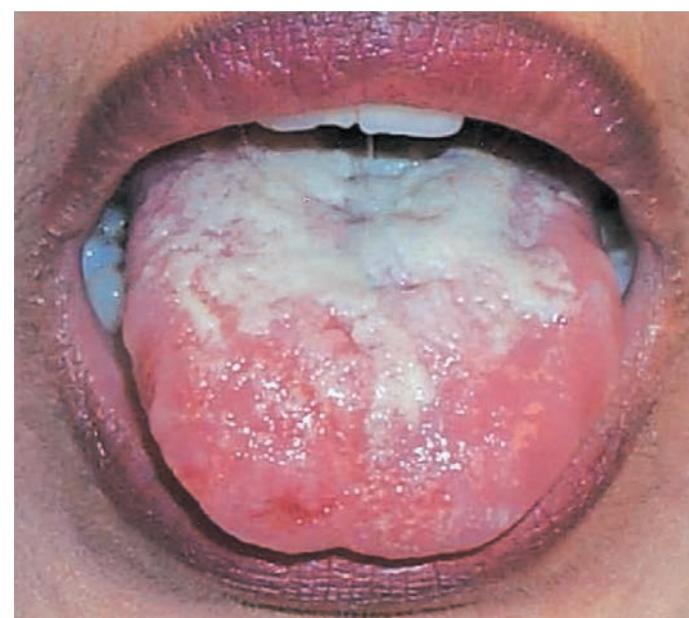
**A****B**

FIGURE 59-1 Pemphigus vulgaris. **A.** Flaccid bullae are easily ruptured, resulting in multiple erosions and crusted plaques. **B.** Involvement of the oral mucosa, which is almost invariable, may present with erosions on the gingiva, buccal mucosa, palate, posterior pharynx, or tongue. (Figure B: Courtesy of Robert Swerlick, MD.)

PV can be life-threatening. Prior to the availability of glucocorticoids, mortality rates ranged from 60% to 90%; the current figure is ~5%. Common causes of morbidity and death are infection and complications of treatment. Bad prognostic factors include advanced age, widespread involvement, and the requirement for high doses of glucocorticoids (with or without other immunosuppressive agents) for control of disease. The course of PV in individual patients is variable and difficult to predict. Some patients experience remission, while others may require long-term treatment or succumb to complications of their disease or its treatment. The mainstay of treatment is systemic glucocorticoids alone or in combination with other immunosuppressive agents. Patients with moderate to severe PV are usually started on prednisone at doses ≤ 1 mg/kg per day (single morning dose). If new lesions continue to appear after 1–2 weeks of treatment, the dose of prednisone may need to be increased and/or combined with another immunosuppressive agent. Among these, rituximab in combination with prednisone often achieves remission (though maintenance therapy may be required to prevent relapse). Other immunosuppressive agents sometimes combined with prednisone to treat PV include azathioprine, mycophenolate mofetil, or cyclophosphamide. Patients with severe, treatment-resistant disease may derive benefit from plasmapheresis (six high-volume exchanges [i.e., 2–3 L per exchange] over ~2 weeks) and/or IV immunoglobulin (IVIg). It is important to bring severe or progressive disease under control quickly in order to lessen the severity and/or duration of this disorder. Increasingly, rituximab and daily glucocorticoids are used early in PV patients to avert the development of advanced and/or treatment-resistant disease.

■ PEMPHIGUS FOLIACEUS

Pemphigus foliaceus (PF) is distinguished from PV by several features. In PF, acantholytic blisters are located high within the epidermis, usually just beneath the stratum corneum. Hence, PF is a more superficial blistering disease than PV. The distribution of lesions in the two disorders is much the same, except that in PF mucous membranes are almost always spared. Patients with PF rarely have intact blisters but rather exhibit shallow erosions associated with erythema, scale, and crust formation. Mild cases of PF can resemble severe seborrheic dermatitis; severe PF may cause extensive exfoliation. Sun exposure (ultraviolet irradiation) may be an aggravating factor.

PF has immunopathologic features in common with PV. Specifically, direct immunofluorescence microscopy of perilesional skin demonstrates IgG on the surface of keratinocytes. Similarly, patients with PF have circulating IgG autoantibodies directed against the surface of keratinocytes. In PF, autoantibodies are directed against Dsg1, a 160-kDa desmosomal cadherin. These autoantibodies can be quantitated by ELISA. As noted for PV, the autoantibody profile in patients with PF (i.e., anti-Dsg1 IgG) and the tissue distribution of this autoantigen (i.e., expression in oral mucosa that is compensated by coexpression of Dsg3) are thought to account for the distribution of lesions in this disease.

Endemic forms of PF are found in south-central rural Brazil, where the disease is known as *fogo salvagem* (FS), as well as in selected sites in Latin America and Tunisia. Endemic PF, like other forms of this disease, is mediated by IgG autoantibodies to Dsg1. Clusters of FS overlap with those of leishmaniasis, a disease transmitted by bites of the sand fly *Lutzomyia longipalis*. Studies have shown that sand fly salivary antigens (specifically, the LJM11 salivary protein) are recognized by IgG autoantibodies from FS patients (as well as by monoclonal antibodies to Dsg1 derived from these patients). The demonstration that mice immunized with LJM11 produce antibodies to Dsg1 suggests that insect bites may deliver salivary antigens, initiate a cross-reactive humoral immune response, and lead to FS in genetically susceptible individuals.

Although pemphigus has been associated with several autoimmune diseases, its association with thymoma and/or myasthenia gravis is particularly notable. To date, >30 cases of thymoma and/or myasthenia gravis have been reported in association with pemphigus, usually with PF. Patients may also develop pemphigus as a consequence of drug exposure; drug-induced pemphigus usually resembles PF rather than PV. Drugs containing a thiol group in their chemical structure (e.g., penicillamine, captopril, enalapril) are most commonly associated with

drug-induced pemphigus. Nonthiol drugs linked to pemphigus include penicillins, cephalosporins, and piroxicam. Some cases of drug-induced pemphigus are durable and require treatment with systemic glucocorticoids and/or immunosuppressive agents.

PF is generally a less severe disease than PV and usually carries a better prognosis. Localized disease can sometimes be treated with topical or intralesional glucocorticoids; more active cases can usually be controlled with systemic glucocorticoids either alone or in combination with other immunosuppressive agents. Patients with severe, treatment-resistant disease may require more aggressive interventions, as described above for patients with PV.

■ PARANEOPLASTIC PEMPHIGUS

Paraneoplastic pemphigus (PNP) is an autoimmune acantholytic mucocutaneous disease associated with an occult or confirmed neoplasm. Patients with PNP typically have painful stomatitis in association with papulosquamous and/or lichenoid eruptions that often progress to blisters. Palm and sole involvement are common in these patients and raise the possibility that prior reports of neoplasia-associated erythema multiforme may have represented unrecognized cases of PNP. Biopsies of lesional skin from these patients show varying combinations of acantholysis, keratinocyte necrosis, and vacuolar-interface dermatitis. Direct immunofluorescence microscopy of a patient's skin shows deposits of IgG and complement on the surface of keratinocytes and (variably) similar immunoreactants in the epidermal basement membrane zone. Patients with PNP have IgG autoantibodies to cytoplasmic proteins that are members of the plakin family (e.g., desmoplakins I and II, bullous pemphigoid antigen [BPAG] 1, enveloplakin, periplakin, and plectin) and to cell-surface proteins that are members of the cadherin family (e.g., Dsg1 and Dsg3). Passive transfer studies have shown that autoantibodies from patients with PNP are pathogenic in animal models.

The predominant neoplasms associated with PNP are non-Hodgkin's lymphoma, chronic lymphocytic leukemia, thymoma, spindle cell tumors, Waldenström's macroglobulinemia, and Castleman's disease; the last-mentioned neoplasm is particularly common among children with PNP. Rare cases of seronegative PNP have been reported in patients with B-cell malignancies previously treated with rituximab. In addition to severe skin lesions, many patients with PNP develop life-threatening bronchiolitis obliterans. PNP is generally resistant to conventional therapies (i.e., those used to treat PV); rarely, a patient's disease may ameliorate or even remit following ablation or removal of underlying neoplasms.

■ BULLOUS PEMPHIGOID

Bullous pemphigoid (BP) is a polymorphic autoimmune subepidermal blistering disease usually seen in the elderly. Initial lesions may consist of urticarial plaques; most patients eventually display tense blisters on either normal-appearing or erythematous skin (Fig. 59-2). The lesions are usually distributed over the lower abdomen, groin, and flexor surface of the extremities; oral mucosal lesions are found in some patients. Pruritus may be nonexistent or severe. As lesions evolve, tense blisters tend to rupture and be replaced by erosions with or without surmounting crust. Nontraumatized blisters heal without scarring. The major histocompatibility complex class II allele HLA-DQ $\beta 1^*0301$ is prevalent in patients with BP. Though most cases occur sporadically, BP can be triggered by medications (e.g., furosemide, dipeptidyl peptidase-4 inhibitors, immune checkpoint inhibitors), ultraviolet light, or ionizing radiation. Several studies have shown that BP is associated with neurologic diseases (e.g., stroke, dementia, Parkinson's disease, and multiple sclerosis).

Biopsies of early lesional skin demonstrate subepidermal blisters and histologic features that roughly correlate with the clinical character of the lesion under study. Lesions on normal-appearing skin generally contain a sparse perivascular leukocytic infiltrate with some eosinophils; conversely, biopsies of inflammatory lesions typically show an eosinophil-rich infiltrate at sites of vesicle formation and in perivasculär areas. In addition to eosinophils, cell-rich lesions also contain mononuclear cells and neutrophils. It is not possible to distinguish



FIGURE 59-2 Bullous pemphigoid with tense vesicles and bullae on erythematous, urticarial bases. (Courtesy of the Yale Resident's Slide Collection; with permission.)

BP from other subepidermal blistering diseases by routine histologic studies alone.

Direct immunofluorescence microscopy of normal-appearing perilesional skin from patients with BP shows linear deposits of IgG and/or C3 in the epidermal basement membrane. The sera of ~70% of these patients contain circulating IgG autoantibodies that bind the epidermal basement membrane of normal human skin in indirect immunofluorescence microscopy. IgG from an even higher percentage of patients reacts with the epidermal side of 1 M NaCl split skin (an alternative immunofluorescence microscopy test substrate used to distinguish circulating IgG autoantibodies to the basement membrane in patients with BP from those in patients with similar, yet different, subepidermal blistering diseases; see below). In BP, circulating autoantibodies recognize 230- and 180-kDa hemidesmosome-associated proteins in basal keratinocytes (i.e., BPAG1 and BPAG2, respectively). Autoantibodies to BPAG2 are thought to deposit *in situ*, activate complement, produce dermal mast-cell degranulation, and generate granulocyte-rich infiltrates that cause tissue damage and blister formation.

BP may persist for months to years, with exacerbations or remissions. Extensive involvement may result in widespread erosions and compromise cutaneous integrity; elderly and/or debilitated patients may die. Local or minimal disease can sometimes be controlled with potent topical glucocorticoids alone; more extensive lesions generally respond to systemic glucocorticoids either alone or in combination with other agents. Adjuncts to systemic glucocorticoids include doxycycline, azathioprine, mycophenolate mofetil, and rituximab.

■ PEMPHIGOID GESTATIONIS

Pemphigoid gestationis (PG), also known as *herpes gestationis*, is a rare, nonviral, subepidermal blistering disease of pregnancy and the puerperium. PG may begin during any trimester of pregnancy or present shortly after delivery. Lesions are usually distributed over the abdomen, trunk, and extremities; mucous membrane lesions are rare. Skin lesions in these patients may be quite polymorphic and consist of erythematous urticarial papules and plaques, vesiculopapules, and/or frank bullae. Lesions are almost always extremely pruritic. Severe exacerbations of PG frequently follow delivery, typically within 24–48 h. PG tends to recur in subsequent pregnancies, often beginning earlier during such gestations. Brief flare-ups of disease may occur with resumption of menses and may develop in patients later exposed to oral contraceptives. Occasionally, infants of affected mothers have transient skin lesions.

Biopsies of early lesional skin show teardrop-shaped subepidermal vesicles forming in dermal papillae in association with an eosinophil-rich leukocytic infiltrate. Differentiation of PG from other subepidermal

bullous diseases by light microscopy is difficult. However, direct immunofluorescence microscopy of perilesional skin from PG patients reveals the immunopathologic hallmark of this disorder: linear deposits of C3 in the epidermal basement membrane. These deposits develop as a consequence of complement activation produced by low-titer IgG anti-basement membrane autoantibodies directed against BPAG2, the same hemidesmosome-associated protein that is targeted by autoantibodies in patients with BP—a subepidermal bullous disease that resembles PG clinically, histologically, and immunopathologically.

The goals of therapy in patients with PG are to prevent the development of new lesions, relieve intense pruritus, and care for erosions at sites of blister formation. Many patients require treatment with moderate doses of daily glucocorticoids (i.e., 20–40 mg of prednisone) at some point in their course. Mild cases (or brief flare-ups) may be controlled by vigorous use of potent topical glucocorticoids. Infants born of mothers with PG appear to be at increased risk of being born slightly premature or “small for dates.” Current evidence suggests that there is no difference in the incidence of uncomplicated live births between PG patients treated with systemic glucocorticoids and those managed more conservatively. If systemic glucocorticoids are administered, newborns are at risk for development of reversible adrenal insufficiency.

■ DERMATITIS HERPETIFORMIS

Dermatitis herpetiformis (DH) is an intensely pruritic, papulovesicular skin disease characterized by lesions symmetrically distributed over extensor surfaces (i.e., elbows, knees, buttocks, back, scalp, and posterior neck) (see Fig. 56-8). Primary lesions in this disorder consist of papules, papulovesicles, or urticarial plaques. Because pruritus is prominent, patients may present with excoriations and crusted papules but no observable primary lesions. Patients sometimes report that their pruritus has a distinctive burning or stinging component; the onset of such local symptoms reliably heralds the development of distinct clinical lesions 12–24 h later. Almost all DH patients have associated, usually subclinical, gluten-sensitive enteropathy (Chap. 325), and >90% express the HLA-B8/DRw3 and HLA-DQw2 haplotypes. DH may present at any age, including in childhood; onset in the second to fourth decades is most common. The disease is typically chronic.

Biopsy of early lesional skin reveals neutrophil-rich infiltrates within dermal papillae. Neutrophils, fibrin, edema, and microvesicle formation at these sites are characteristic of early disease. Older lesions may demonstrate nonspecific features of a subepidermal bulla or an excoriated papule. Because the clinical and histologic features of this disease can be variable and resemble those of other subepidermal blistering disorders, the diagnosis is confirmed by direct immunofluorescence microscopy of normal-appearing perilesional skin. Such studies demonstrate granular deposits of IgA (with or without complement components) in the papillary dermis and along the epidermal basement membrane zone. IgA deposits in the skin are unaffected by control of disease with medication; however, these immunoreactants diminish in intensity or disappear in patients maintained for long periods on a strict gluten-free diet (see below). Patients with DH have granular deposits of IgA in their epidermal basement membrane zone and should be distinguished from individuals with linear IgA deposits at this site (see below).

Although most DH patients do not report overt gastrointestinal symptoms or have laboratory evidence of malabsorption, biopsies of the small bowel usually reveal blunting of intestinal villi and a lymphocytic infiltrate in the lamina propria. As is true for patients with celiac disease, this gastrointestinal abnormality can be reversed by a gluten-free diet. Moreover, if maintained, this diet alone may control the skin disease and eventuate in clearance of IgA deposits from these patients’ epidermal basement membrane zones. Subsequent gluten exposure in such patients alters the morphology of their small bowel, elicits a flare-up of their skin disease, and is associated with the reappearance of IgA in their epidermal basement membrane zones. As in patients with celiac disease, dietary gluten sensitivity in patients with DH is associated with IgA anti-endomysial autoantibodies that target tissue transglutaminase. Studies indicate that patients with DH also

have high-avidity IgA autoantibodies to epidermal transglutaminase and that the latter is co-localized with granular deposits of IgA in the papillary dermis of DH patients. Patients with DH also have an increased incidence of thyroid abnormalities, achlorhydria, atrophic gastritis, and autoantibodies to gastric parietal cells. These associations likely relate to the high frequency of the HLA-B8/DRw3 haplotype in these patients, since this marker is commonly linked to autoimmune disorders. The mainstay of treatment of DH is dapsone, a sulfone. Patients respond rapidly (24–48 h) to dapsone, but require careful pre-treatment evaluation (e.g., screening for glucose-6-phosphate dehydrogenase deficiency) and close follow-up to ensure that complications are avoided or controlled. All patients taking dapsone at >100 mg/d will have some hemolysis and methemoglobinemia, which are expected pharmacologic side effects of this agent. Gluten restriction can control DH and lessen dapsone requirements; this diet must rigidly exclude gluten to be of maximal benefit. Many months of dietary restriction may be necessary before a beneficial result is achieved. Good dietary counseling by a trained dietitian is essential.

■ LINEAR IGA DISEASE

Linear IgA disease, once considered a variant form of DH, is actually a separate and distinct entity. Clinically, patients with linear IgA disease may resemble individuals with DH, BP, or other subepidermal blistering diseases. Lesions typically consist of papulovesicles, bullae, and/or urticarial plaques that develop predominantly on central or flexural sites. Oral mucosal involvement occurs in some patients. Severe pruritus resembles that seen in patients with DH. Patients with linear IgA disease do not have an increased frequency of the HLA-B8/DRw3 haplotype or an associated enteropathy and therefore are not candidates for treatment with a gluten-free diet.

Histologic alterations in early lesions may be virtually indistinguishable from those in DH. However, direct immunofluorescence microscopy of normal-appearing perilesional skin reveals a linear band of IgA (and often C3) in the epidermal basement membrane zone. Most patients with linear IgA disease have circulating IgA anti-basement membrane autoantibodies directed against neopeptides in the proteolytically processed extracellular domain of BPAG2. These patients generally respond to treatment with dapsone (50–200 mg/d) alone or in combination with low daily doses of prednisone.

■ EPIDERMOLYSIS BULLOSA ACQUISITA

Epidermolysis bullosa acquisita (EBA) is a rare, noninherited, polymorphic, chronic, subepidermal blistering disease. ([The inherited form is discussed in Chap. 413.](#)) Patients with classic or noninflammatory EBA have blisters on noninflamed skin, atrophic scars, milia, nail dystrophy, hair loss, and oral lesions. Because lesions generally occur at sites exposed to minor trauma, classic EBA is considered a mechanobullous disease. Other patients with EBA have widespread inflammatory scarring and bullous lesions that resemble severe BP. Inflammatory EBA may evolve into the classic, noninflammatory form of this disease. Rarely, patients present with lesions that predominate on mucous membranes. The HLA-DR2 haplotype is found with increased frequency in EBA patients. Studies suggest that EBA is sometimes associated with inflammatory bowel disease (especially Crohn's disease).

The histology of lesional skin varies with the character of the lesion being studied. Noninflammatory bullae are subepidermal, feature a sparse leukocytic infiltrate, and resemble the lesions in patients with porphyria cutanea tarda. Inflammatory lesions consist of neutrophil-rich subepidermal blisters. EBA patients have continuous deposits of IgG (and frequently C3) in a linear pattern within the epidermal basement membrane zone. Ultrastructurally, these immunoreactants are found in the sublamina densa region in association with anchoring fibrils. Approximately 50% of EBA patients have demonstrable circulating IgG anti-basement membrane autoantibodies directed against type VII collagen—the collagen species that makes up anchoring fibrils. Such IgG autoantibodies bind the dermal side of 1 M NaCl split skin (in contrast to IgG autoantibodies in patients with BP). Studies have shown that passive transfer of experimental or patient IgG against type

VII collagen can produce lesions in mice that clinically, histologically, and immunopathologically resemble those in patients with EBA.

Treatment of EBA is generally unsatisfactory. Some patients with inflammatory EBA may respond to systemic glucocorticoids, either alone or in combination with immunosuppressive agents. Other patients (especially those with neutrophil-rich inflammatory lesions) may respond to dapsone. The chronic, noninflammatory form of EBA is largely resistant to treatment, although some patients may respond to prednisone in combination with rituximab, cyclosporine, mycophenolate mofetil, azathioprine, or IVIg.

■ MUCOUS MEMBRANE PEMPHIGOID

Mucous membrane pemphigoid (MMP) is a rare, acquired, subepithelial immunobullous disease characterized by erosive lesions of mucous membranes and skin that result in scarring of at least some sites of involvement. Common sites include the oral mucosa (especially the gingiva) and conjunctiva; other sites that may be affected include the nasopharyngeal, laryngeal, esophageal, and anogenital mucosa. Skin lesions (present in about one-third of patients) tend to predominate on the scalp, face, and upper trunk and generally consist of a few scattered erosions or tense blisters on an erythematous or urticarial base. MMP is typically a chronic and progressive disorder. Serious complications may arise as a consequence of ocular, laryngeal, esophageal, or anogenital lesions. Erosive conjunctivitis may result in shortened fornices, symblepharon, ankyloblepharon, entropion, corneal opacities, and (in severe cases) blindness. Similarly, erosive lesions of the larynx may cause hoarseness, pain, and tissue loss that, if unrecognized and untreated, may eventuate in complete destruction of the airway. Esophageal lesions may result in stenosis and/or strictures that could place patients at risk for aspiration. Strictures may also complicate anogenital involvement.

Biopsies of lesional tissue generally show subepithelial vesiculobullae and a mononuclear leukocytic infiltrate. Neutrophils and eosinophils may be seen in biopsies of early lesions; older lesions may demonstrate a scant leukocytic infiltrate and fibrosis. Direct immunofluorescence microscopy of perilesional tissue typically reveals deposits of IgG, IgA, and/or C3 in the epidermal basement membrane. Because many patients with MMP exhibit no evidence of circulating anti-basement membrane autoantibodies, testing of perilesional skin is important diagnostically. Although MMP was once thought to be a single nosologic entity, it is now largely regarded as a disease phenotype that may develop as a consequence of an autoimmune reaction to a variety of molecules in the epidermal basement membrane (e.g., BPAG2, laminin-332, type VII collagen, $\alpha_6\beta_4$ integrin) and other antigens yet to be completely defined. Studies suggest that MMP patients with autoantibodies to laminin-332 have an increased relative risk for cancer. Treatment of MMP is largely dependent upon the sites of involvement. Due to potentially severe complications, patients with ocular, laryngeal, esophageal, and/or anogenital involvement require aggressive systemic treatment with dapsone, prednisone, or the latter in combination with another immunosuppressive agent (e.g., rituximab, azathioprine, mycophenolate mofetil, or cyclophosphamide), or IVIg. Less threatening forms of the disease may be managed with topical or intralesional glucocorticoids.

AUTOIMMUNE SYSTEMIC DISEASES WITH PROMINENT CUTANEOUS FEATURES

■ DERMATOMYOSITIS

The cutaneous manifestations of dermatomyositis ([Chap. 365](#)) are often distinctive but at times may resemble those of systemic lupus erythematosus (SLE) ([Chap. 356](#)), scleroderma ([Chap. 360](#)), or other overlapping connective tissue diseases ([Chap. 360](#)). The extent and severity of cutaneous disease may or may not correlate with the extent and severity of the myositis. The cutaneous manifestations of dermatomyositis are similar, whether the disease appears in children or in the elderly, except that calcification of subcutaneous tissue is a common late sequela in childhood dermatomyositis. Dermatomyositis may be associated with interstitial lung disease or cancer.



FIGURE 59-3 Dermatomyositis. Periorbital violaceous erythema characterizes the classic heliotrope rash. (Courtesy of James Krell, MD; with permission.)

The cutaneous signs of dermatomyositis may precede or follow the development of myositis by weeks to years. Cases lacking muscle involvement (i.e., *dermatomyositis sine myositis* or *amyopathic dermatomyositis*) have also been reported. The most common manifestation is a purple-red discoloration of the upper eyelids, sometimes associated with scaling (“heliotrope” erythema; **Fig. 59-3**) and periorbital edema. Erythema on the cheeks and nose in a “butterfly” distribution may resemble the malar eruption of SLE. Erythematous or violaceous thin, scaly plaques are common on the upper trunk and neck (shawl sign), the scalp, lateral aspects of the thighs (holster sign), and the extensor surfaces of the forearms and hands (tendon streaking). Approximately one-third of patients have violaceous, flat-topped papules over the dorsal interphalangeal joints that are pathognomonic of dermatomyositis (Gottron’s papules) (**Fig. 59-4**). Thin violaceous papules and plaques on the elbows and knees of patients with dermatomyositis are referred to as *Gottron’s sign*. These lesions can be contrasted with the erythema and scaling on the dorsum of the fingers that spares the skin over the interphalangeal joints of some SLE patients. Periungual telangiectasias and edema may be prominent in patients with dermatomyositis. Other patients, particularly those with long-standing disease, develop areas of hypopigmentation, hyperpigmentation, mild atrophy, and telangiectasia known as *poikiloderma*. Poikiloderma is rare in both SLE and scleroderma and thus can serve as a clinical sign that distinguishes dermatomyositis from these two diseases. Cutaneous changes may be similar in dermatomyositis and various overlap syndromes where thickening and binding down of the skin of the hands (*sclerodactyly*)



FIGURE 59-4 Gottron’s papules. Dermatomyositis often involves the hands as erythematous flat-topped papules over the knuckles. Periungual telangiectasias are also evident.

as well as Raynaud’s phenomenon can be seen. However, the presence of severe muscle disease, Gottron’s papules, heliotrope erythema, and poikiloderma serves to distinguish patients with dermatomyositis. Skin biopsy of the erythematous, scaling lesions of dermatomyositis may reveal only mild nonspecific inflammation, but sometimes may show changes indistinguishable from those found in cutaneous lupus erythematosus (LE), including epidermal atrophy, hydropic degeneration of basal keratinocytes, and dermal changes consisting of interstitial mucin deposition and a mild mononuclear cell perivascular infiltrate. Direct immunofluorescence microscopy of lesional skin is usually negative, although granular deposits of immunoglobulin(s) and complement in the epidermal basement membrane zone have been described in some patients. Treatment should be stratified based on the relative severity of disease. Topical treatments include glucocorticoids, sunscreens, and aggressive photoprotective measures. Treatment of systemic disease includes antimalarials (though some patients may develop a drug eruption upon initiation of therapy) or systemic glucocorticoids in conjunction with methotrexate, mycophenolate mofetil, azathioprine, rituximab, or IVIg.

LUPUS ERYTHEMATOSUS

The cutaneous manifestations of LE (**Chap. 356**) can be divided into acute, subacute, and chronic types. *Acute cutaneous LE* is characterized by erythema of the nose and malar eminences in a “butterfly” distribution (**Fig. 59-5A**). The erythema is often sudden in onset, accompanied



A



FIGURE 59-5 Acute cutaneous lupus erythematosus (LE). **A.** Acute cutaneous LE on the face, showing prominent, scaly, malar erythema. Involvement of other sun-exposed sites is also common. **B.** Acute cutaneous LE on the upper chest, demonstrating brightly erythematous and slightly edematous papules and plaques. (Source: B, Courtesy of Robert Sverlick, MD; with permission.)

by edema and fine scale, and correlated with systemic involvement. Patients may have widespread involvement of the face as well as erythema and scaling of the extensor surfaces of the extremities and upper chest (Fig. 59-5B). These acute lesions, while sometimes evanescent, usually last for days and are often associated with exacerbations of systemic disease. Skin biopsy of acute lesions typically shows hydropic degeneration of basal keratinocytes, dermal edema, and (in some cases) a sparse perivascular infiltrate of mononuclear cells in the upper dermis as well as dermal mucin. Direct immunofluorescence microscopy of lesional skin frequently reveals deposits of immunoglobulin(s) and complement in the epidermal basement membrane zone. Treatment of cutaneous disease includes topical glucocorticoids, aggressive photoprotection, antimalarials, and control of systemic disease. Treatment of systemic disease associated with acute cutaneous LE includes systemic glucocorticoids in conjunction with other immunosuppressive agents.

Subacute cutaneous lupus erythematosus (SCLE) is characterized by a widespread photosensitive, nonscarring eruption. In most patients, renal and central nervous system involvement is mild or absent. SCLE may present as a papulosquamous eruption that resembles psoriasis or as annular polycyclic lesions. In the papulosquamous form, discrete erythematous papules arise on the back, chest, shoulders, extensor surfaces of the arms, and dorsum of the hands; lesions are uncommon on the central face and the flexor surfaces of the arms as well as below the waist. These slightly scaling papules tend to merge into plaques. The annular form involves the same areas and presents with erythematous papules that evolve into oval, circular, or polycyclic lesions. The lesions of SCLE are more widespread but have less tendency for scarring than lesions of discoid LE. In many patients with SCLE, drugs (e.g., hydrochlorothiazide, calcium channel blockers, antifungals, proton pump inhibitors) may induce or exacerbate disease. Skin biopsy typically reveals epidermal changes that include atrophy, hydropic degeneration of basal keratinocytes, and apoptosis accompanied by an infiltrate of mononuclear cells in the upper dermis. Direct immunofluorescence microscopy of lesional skin reveals deposits of immunoglobulin(s) in the epidermal basement membrane zone in about one-half of these cases. A particulate pattern of IgG deposition throughout the epidermis has been associated with SCLE. Most SCLE patients have anti-Ro autoantibodies. Local therapy alone is usually unsuccessful. Most patients require treatment with aminoquinoline antimalarial drugs. Low-dose therapy with oral glucocorticoids is sometimes necessary. Photoprotective measures against both ultraviolet B and ultraviolet A wavelengths are very important.

Chronic cutaneous LE has multiple subtypes; *discoid LE* (DLE) is the most common. DLE is characterized by discrete lesions, most often found on the face, scalp, and/or external ears. The lesions are erythematous papules or plaques with a thick, adherent scale that occludes hair follicles (follicular plugging). When the scale is removed, its underside shows small excrescences that correlate with the openings of hair follicles (so-called “carpet tacking”), a finding relatively specific for DLE. Long-standing lesions develop central atrophy, scarring, and hypopigmentation but frequently have erythematous, sometimes raised borders (Fig. 59-6). These lesions persist for years and tend to expand slowly. Up to 20% of patients with DLE eventually meet the American College of Rheumatology criteria for SLE. Typical discoid lesions are frequently seen in patients with SLE. Biopsy of DLE lesions shows hyperkeratosis, follicular plugging, atrophy of the epidermis, hydropic degeneration of basal keratinocytes, thickening of the epidermal basement membrane zone, and a mononuclear cell infiltrate adjacent to epidermal, adnexal, and microvascular basement membranes. Direct immunofluorescence microscopy demonstrates immunoglobulin(s) and complement deposits at the basement membrane zone in ~90% of cases. Treatment is focused on control of local cutaneous disease and consists mainly of photoprotection and topical or intralesional glucocorticoids. If local therapy is ineffective, use of aminoquinoline antimalarial agents may be indicated.

■ SCLERODERMA AND MORPHEA

The skin changes of scleroderma (Chap. 360) may be limited or diffuse. In both instances, disease usually begins on the fingers, hands, toes, feet, and face, with episodes of recurrent nonpitting edema.



FIGURE 59-6 Discoid lupus erythematosus (DLE). Violaceous, hyperpigmented, atrophic plaques, follicular plugging, and scarring are typical features of DLE.

Sclerosis of the skin commences distally on the fingers (sclerodactyly) and spreads proximally, usually accompanied by resorption of bone of the fingertips, which may have punched out ulcers, stellate scars, or areas of hemorrhage (Fig. 59-7). The fingers may shrink and become sausage-shaped, and, because the fingernails are usually unaffected, they may curve over the end of the fingertips. Periungual telangiectasias are usually present, but periungual erythema is rare. In diffuse disease, the extremities show contractures and calcinosis cutis; facial involvement includes a smooth, unwrinkled brow, taut skin over the nose, shrinkage of tissue around the mouth, and perioral radial furrowing (Fig. 59-8). Matlike telangiectasias are often present, particularly on the face and hands. Involved skin feels indurated, smooth, and bound to underlying structures; hyper- and hypopigmentation are common as well. *Raynaud's phenomenon* (i.e., cold-induced blanching, cyanosis, and reactive hyperemia) is documented in almost all patients and can precede development of scleroderma by many years. The combination of calcinosis cutis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasias has been termed as the *CREST syndrome*. Anti-centromere autoantibodies have been reported in a very high percentage of patients with CREST syndrome but in only a small minority of patients with scleroderma. Skin biopsy reveals thickening of the dermis, homogenization of collagen bundles, atrophic pilosebaceous and eccrine glands, and a sparse mononuclear cell infiltrate in the dermis and subcutaneous fat. Direct immunofluorescence microscopy of lesional skin is usually negative. Treatments for



FIGURE 59-7 Scleroderma showing acral sclerosis and focal digital ulcers.



FIGURE 59-8 Scleroderma often eventuates in development of an expressionless, masklike facies.

cutaneous disease include emollients, antipruritics, and phototherapy (UVA1 [ultraviolet A1 irradiation] or PUVA [psoralens + ultraviolet A irradiation]). Treatment of systemic disease includes vascular modifying agents, immunosuppressives, and antifibrotics.

Morphea is characterized by localized thickening and sclerosis of skin; it dominates on the trunk. This disorder may affect children or adults. Morphea begins as erythematous or flesh-colored plaques that become sclerotic, develop central hypopigmentation, and have an erythematous border. In most cases, patients have one or a few lesions, and the disease is termed *circumscribed morphea*. In some patients, widespread cutaneous lesions may occur without systemic involvement (*generalized morphea*). Many adults with generalized morphea have concomitant rheumatic or other autoimmune disorders. Skin biopsy of morphea is generally indistinguishable from that of scleroderma. Scleroderma and morphea are usually quite resistant to therapy. For this reason, physical therapy to prevent joint contractures and to maintain function is employed and is often helpful. Treatment options for early, rapidly progressive disease include phototherapy (UVA1 or PUVA) or methotrexate alone or in combination with daily glucocorticoids.

Diffuse fasciitis with eosinophilia is a clinical entity that can sometimes be confused with scleroderma. There is usually a sudden onset of swelling, induration, and erythema of the extremities, frequently following significant physical exertion, initiation of hemodialysis, exposure to certain medications, or other triggers. The proximal portions of the extremities (upper arms, forearms, thighs, calves) are more often involved than are the hands and feet. While the skin is indurated, it usually displays a woody, dimpled, or “pseudocellulite” appearance rather than being bound down as in scleroderma; contractures may occur early secondary to fascial involvement. The latter may also cause muscle groups to be separated and veins to appear depressed (i.e., the “groove sign”). These skin findings are accompanied by peripheral-blood eosinophilia, increased erythrocyte sedimentation rate, and sometimes hypergammaglobulinemia. Deep biopsy of affected areas of skin reveals inflammation and thickening of the deep fascia overlying muscle. An inflammatory infiltrate composed of eosinophils and mononuclear cells is usually found. Patients with eosinophilic fasciitis appear to be at increased risk for developing bone marrow failure or other hematologic abnormalities. While the ultimate course of eosinophilic fasciitis is variable, most patients respond favorably to treatment with prednisone. Relapses may occur and require treatment with prednisone in combination with other immunosuppressive or immunomodulatory agents.

The *eosinophilia-myalgia syndrome*, a disorder with epidemic numbers of cases reported in 1989 and linked to ingestion of L-tryptophan manufactured by a single company in Japan, is a multisystem disorder characterized by debilitating myalgias and absolute eosinophilia in association with varying combinations of arthralgias, pulmonary

symptoms, and peripheral edema. In a later phase (3–6 months after initial symptoms), these patients often develop localized scleroderma-like skin changes, weight loss, and/or neuropathy (**Chap. 360**).

FURTHER READING

- BOLOGNA JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.
- HAMMERS CM, STANLEY JR: Mechanisms of disease: Pemphigus and bullous pemphigoid. *Annu Rev Pathol* 11:175, 2016.
- KANG S et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 9th ed. New York, McGraw-Hill, 2019.
- SCHMIDT E, ZILLIKENS D: Pemphigoid diseases. *Lancet* 381:320, 2013.

60

Cutaneous Drug Reactions

Robert G. Micheletti, Misha Rosenbach,
Bruce U. Wintrob, Kanade Shinkai



Cutaneous reactions are the most frequent adverse reactions to medications, representing 10–15% of reported adverse drug reactions. Most are benign, but a few can be life threatening. Prompt recognition of severe reactions, drug withdrawal, and appropriate therapeutic interventions can minimize toxicity. This chapter focuses on adverse cutaneous reactions to systemic medications; it covers their incidence, patterns, and pathogenesis, and provides some practical guidelines on treatment, assessment of causality, and future use of drugs.

USE OF PRESCRIPTION DRUGS IN THE UNITED STATES

In the United States, more than 4 billion prescriptions for >60,000 drug products are dispensed annually. Hospital inpatients alone annually receive about 120 million courses of drug therapy, and half of adult Americans receive prescription drugs on a regular outpatient basis. Adverse effects of a prescription medication may result in 4.5 million urgent or emergency care visits and over 7000 deaths each year in the United States. Many patients use over-the-counter medicines that may cause adverse cutaneous reactions.

INCIDENCE OF CUTANEOUS REACTIONS

Several recent prospective studies reported that acute cutaneous reactions to drugs affect between 2.2 and 10 per 1000 hospitalized patients. Reactions usually occur a few days to 4 weeks after initiation of therapy.

In a series of 48,005 inpatients over a 20-year period, morbilliform rash (91%) and urticaria (6%) were the most frequent skin reactions, and antimicrobials, radiocontrast, and nonsteroidal anti-inflammatory drugs (NSAIDs) were the most common drug associations. Severe hypersensitivity reactions to medications have been reported to occur in between 1 in 1000 to 2 per million users, depending on the reaction type. Although rare, severe cutaneous reactions to drugs have an important impact on health because of significant sequelae; in addition, they may require hospitalization, increase the duration of hospital stay, or be life threatening. Some populations are at increased risk of drug reactions, including elderly patients, patients with autoimmune disease, hematopoietic stem cell transplant recipients, and those with acute Epstein-Barr virus (EBV) or human immunodeficiency virus (HIV) infection. The pathophysiology underlying this association is unknown but may be related to immune dysregulation. Individuals with advanced HIV disease (e.g., CD4 T lymphocyte count <200 cells/ μ L) have a 40- to 50-fold increased risk of adverse reactions to sulfamethoxazole (**Chap. 202**) and increased risk of severe hypersensitivity reactions.

In addition to acute eruptions, a variety of skin diseases can be induced or exacerbated by prolonged use of drugs (e.g., pruritus, pigmentation, nail or hair disorders, psoriasis, bullous pemphigoid, photosensitivity, and even cutaneous neoplasms). These drug reactions are not frequent; however, neither their incidence nor their impact on public health has been evaluated.

PATHOGENESIS OF DRUG REACTIONS

Adverse cutaneous responses to drugs can arise as a result of immunologic or nonimmunologic mechanisms.

NONIMMUNOLOGIC DRUG REACTIONS

Examples of nonimmunologic drug reactions are pigmentary changes due to dermal accumulation of medications or their metabolites, alteration of hair follicles by antimetabolites and signaling inhibitors, and lipodystrophy associated with metabolic effects of anti-HIV medications. These side effects are predictable and sometimes can be prevented.

IMMUNOLOGIC DRUG REACTIONS

Evidence suggests an immunologic basis for most acute drug eruptions. Drug reactions may result from immediate release of preformed mediators (e.g., urticaria, anaphylaxis), antibody-mediated reactions, immune complex deposition, and antigen-specific responses. Drug-specific CD4+ and CD8+ T-cell clones can be derived from the blood or from skin lesions of patients with a variety of drug allergies, strongly suggesting that these T cells mediate drug allergy in an antigen-specific manner. Drug presentation to T cells is major histocompatibility complex (MHC)-restricted and likely involves drug-peptide complex recognition by specific T-cell receptors (TCRs).

Once a drug has induced an immune response, the phenotype of the reaction is determined by the nature of effectors: cytotoxic (CD8+) T cells in blistering and certain hypersensitivity reactions, chemokines for reactions mediated by neutrophils or eosinophils, and B cell collaboration for production of specific antibodies for urticarial reactions. Immunologic reactions have recently been classified into further subtypes that provide a useful framework for designating adverse drug reactions based on involvement of specific immune pathways (Table 60-1).

Immediate Reactions Immediate reactions depend on the release of mediators of inflammation by tissue mast cells or circulating basophils. These mediators include histamine, leukotrienes, prostaglandins, bradykinins, platelet-activating factor, enzymes, and proteoglycans. Drugs can trigger mediator release either directly ("anaphylactoid" reaction) or through IgE-specific antibodies. These reactions usually manifest in the skin and gastrointestinal, respiratory, and cardiovascular systems (Chap. 353). Primary symptoms and signs include pruritus, urticaria, nausea, vomiting, abdominal cramps, bronchospasm, laryngeal edema, and, occasionally, anaphylactic shock with hypotension and death. They occur within minutes of drug exposure. NSAIDs, including aspirin, and radiocontrast media are frequent causes of direct mast cell degranulation or anaphylactoid reactions, which can occur on first exposure. Penicillins and muscle relaxants used in general anesthesia are the most frequent causes of IgE-dependent reactions to drugs, which require prior sensitization. Release of mediators is triggered when polyvalent drug protein conjugates cross-link IgE molecules fixed to sensitized cells. Certain routes of administration favor different clinical patterns (e.g., gastrointestinal effects from oral route, circulatory effects from intravenous route).

Immune Complex-Dependent Reactions Serum sickness is produced by tissue deposition of circulating immune complexes with consumption of complement. It is characterized by fever, arthritis, nephritis, neuritis, edema, and an urticarial, papular, or purpuric rash (Chap. 363). First described following administration of nonhuman sera, it currently occurs in the setting of monoclonal antibodies and similar medications. In classic serum sickness, symptoms develop 6 or more days after drug exposure, the latent period representing the time needed to synthesize antibody. Vasculitis, a relatively rare complication

TABLE 60-1 Classification of Adverse Drug Reactions Based on Immune Pathway

Type	Key Pathway	Key Immune Mediators	Adverse Drug Reaction Type
Type I	IgE	IgE	Urticaria, angioedema, anaphylaxis
Type II	IgG-mediated cytotoxicity	IgG	Drug-induced hemolysis, thrombocytopenia (e.g., penicillin)
Type III	Immune complex	IgG + antigen	Vasculitis, serum sickness, drug-induced lupus
Type IVa	T lymphocyte-mediated macrophage inflammation	IFN- γ , TNF- α , T _H 1 cells	Tuberculin skin test, contact dermatitis
Type IVb	T lymphocyte-mediated eosinophil inflammation	IL-4, IL-5, IL-13, T _H 2 cells, Eosinophils	DIHS Morbilliform eruption
Type IVc	T lymphocyte-mediated cytotoxic T lymphocyte inflammation	Cytotoxic T lymphocytes, Granzyme, Perforin, Granulysin (SJS/TEN) only	SJS/TEN Morbilliform eruption
Type IVd	T lymphocyte-mediated neutrophil inflammation	CXCL8, IL-17, GM-CSF, Neutrophils	AGEP

Abbreviations: AGEPE, acute generalized exanthematous pustulosis; DIHS, drug-induced hypersensitivity syndrome; GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN, interferon; IL, interleukin; SJS, Stevens-Johnson syndrome; TEN, toxic epidermal necrolysis; TNF, tumor necrosis factor.

of drugs, may also be a result of immune complex deposition (Chap. 363). Penicillin, cefaclor, amoxicillin, trimethoprim/sulfamethoxazole, and monoclonal antibodies such as infliximab, rituximab, and omalizumab may be associated with clinically similar "serum sickness-like" reactions (SSLR). The mechanism of this reaction is unknown but is unrelated to immune complex formation and complement activation, and systemic involvement is rare. Whereas serum sickness most commonly occurs in adults, SSLR is more frequently observed in children.

Delayed Hypersensitivity While not completely understood, delayed hypersensitivity directed by drug-specific T cells is an important mechanism underlying the most common drug eruptions, that is, morbilliform eruptions, and also rare and severe forms such as drug-induced hypersensitivity syndrome (DIHS) (also known as drug rash with eosinophilia and systemic symptoms [DRESS]), acute generalized exanthematous pustulosis (AGEP), Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN) (Table 60-1). Drug-specific T cells have been detected in these types of drug eruptions. In TEN, skin lesions contain T lymphocytes reactive to autologous lymphocytes and keratinocytes in a drug-specific, human leukocyte antigen (HLA)-restricted, and perforin/granzyme-mediated pathway. In the case of carbamazepine, studies have identified cytotoxic T lymphocytes (CTLs) reactive to carbamazepine that use highly restricted V-alpha and V-beta TCR repertoires in patients with carbamazepine hypersensitivity that are not found in carbamazepine-tolerant individuals.

The mechanism(s) by which medications result in T-cell activation is unknown. Two hypotheses prevail: first, that the antigens driving these reactions may be the native drug itself or components of the drug covalently complexed with endogenous proteins, presented in association with HLA molecules to T cells through the classic antigen presentation pathway or, alternatively, through direct interaction of the drug/metabolite with the TCR or peptide-loaded HLA (e.g., the pharmacologic interaction of drugs with immune receptors, or p-i hypothesis). Recent x-ray crystallography data characterizing binding between

specific HLA molecules to drugs known to cause hypersensitivity reactions demonstrate unique alterations to the MHC peptide-binding groove, suggesting a molecular basis for T-cell activation in the development of hypersensitivity reactions.

■ GENETIC FACTORS AND CUTANEOUS DRUG REACTIONS

 Genetic determinants may predispose individuals to severe drug reactions by affecting either drug metabolism or immune responses to drugs. Polymorphisms in cytochrome P450 enzymes, drug acetylation, methylation (such as thiopurine methyltransferase activity and azathioprine), and other forms of metabolism (such as glucose-6-phosphate dehydrogenase and dapsone) may increase susceptibility to drug toxicity or underdosing and increase risk for medication interactions, highlighting a role for differential pharmacokinetic or pharmacodynamic effects. The value of routine screening of P450 enzymes for prediction of cutaneous reactions has not been determined, though its cost-effectiveness in certain populations (e.g., patients with seizure disorder, depression) as well as patients considering specific therapies (e.g., tamoxifen, warfarin) has been suggested.

Associations between drug hypersensitivities and HLA haplotypes suggest a key role for immune mechanisms, especially those leading to skin involvement. Hypersensitivity to the anti-HIV medication abacavir is strongly associated with HLA-B*57:01 ([Chap. 202](#)). In Taiwan, within a homogeneous Han Chinese population, a strong association was observed between SJS/TEN (but not DIHS) related to carbamazepine and HLA-B*15:02. In the same population, a strong association was found between HLA-B*58:01 and SJS, TEN, or DIHS related to allopurinol. These associations are drug and phenotype specific; that is, HLA-specific T cell stimulation by medications leads to distinct reactions. However, while this genetic association is strong, it is not sufficient to cause severe drug hypersensitivity reactions.

■ GLOBAL CONSIDERATIONS

Recognition of HLA associations with drug hypersensitivity has resulted in recommendations to screen high-risk populations. Genetic screening for HLA-B*57:01 to prevent abacavir hypersensitivity, which carries a 100% negative predictive value when patch test confirmed and 55% positive predictive value generalizable across races, is becoming the clinical standard of care worldwide (number needed to treat = 13). The U.S. Food and Drug Administration has recommended HLA-B*15:02 screening of Asian individuals prior to a new prescription of carbamazepine. The American College of Rheumatology has recommended HLA-B*58:01 screening of Han Chinese patients prescribed allopurinol. To date, screening for a single HLA (but not multiple HLA haplotypes) in specific populations has been determined to be cost-effective (e.g., HLA-B*1301 screening in Chinese patients with leprosy treated with dapsone). Genetic testing for specific HLA haplotypes and functional screening for TCR repertoire to identify patients at risk is becoming more widely available and heralds the era of personalized medicine and pharmacogenomics.

CLINICAL PRESENTATION OF CUTANEOUS DRUG REACTIONS

■ NONIMMUNE CUTANEOUS REACTIONS

Exacerbation or Induction of Dermatologic Diseases A variety of drugs can exacerbate preexisting diseases or induce—or unmask—a disease that may or may not disappear after withdrawal of the inducing medication. For example, NSAIDs, lithium, beta blockers, tumor necrosis factor (TNF) antagonists, interferon (IFN) α , and angiotensin-converting enzyme (ACE) inhibitors can exacerbate plaque psoriasis, whereas antimalarials and withdrawal of systemic glucocorticoids can worsen pustular psoriasis. The situation of TNF- α inhibitors is unusual, as this class of medications is used to treat psoriasis; however, they may induce psoriasis (especially palmoplantar) in patients being treated for other conditions. Acne may be induced by glucocorticoids, androgens, lithium, and antidepressants. Follicular papular or pustular eruptions of the face and trunk resembling

acne frequently occur with epidermal growth factor receptor (EGFR) antagonists, mitogen-activated protein kinase (MEK) inhibitors, and other targeted inhibitors. With EGFR antagonists, the severity of the eruption correlates with a better anticancer effect. This rash is typically responsive to and prevented by tetracycline antibiotics.

Several medications induce or exacerbate autoimmune disease. Checkpoint inhibitors induce a wide array of systemic autoimmune reactions, including in skin. Interleukin (IL) 2, IFN- α , and anti-TNF- α are associated with new-onset systemic lupus erythematosus (SLE). Drug-induced lupus is classically marked by antinuclear and antihistone antibodies and, in some cases, anti-double-stranded DNA (D-penicillamine, anti-TNF- α) or perinuclear antineutrophil cytoplasmic antibodies (p-ANCA) (minocycline). Subacute cutaneous lupus erythematosus (SCLE) can be induced by a growing list of drugs, including thiazide diuretics, proton pump inhibitors, TNF inhibitors, terbinafine, and minocycline. Drug-induced dermatomyositis may rarely occur with TNF inhibitors or capecitabine; hydroxyurea can induce skin findings of dermatomyositis. IFN and TNF inhibitors, as well as checkpoint inhibitors, can induce granulomatous disease and sarcoidosis. Autoimmune blistering diseases may be drug induced as well: pemphigus by D-penicillamine and ACE inhibitors; bullous pemphigoid by DPP4 inhibitors, furosemide, and PD-1 inhibitors; and linear IgA bullous dermatosis by vancomycin. Other medications may cause highly specific cutaneous reactions. Gadolinium contrast has been associated with nephrogenic systemic fibrosis, a condition of sclerosing skin with rare internal organ involvement; advanced renal compromise may be an important risk factor. Granulocyte colony-stimulating factor, azacitidine, all-trans-retinoic acid, the *FLT3* inhibitor class of drugs, and rarely levamisole-contaminated cocaine may induce neutrophilic dermatoses. In this setting, the hypothesis that a drug may be responsible should always be considered, even after the treatment is complete. In addition, reactions may develop in cases of long-term medication therapy due to changes in dosing or host metabolism. Resolution of the cutaneous reaction may be delayed upon discontinuation of the medication.

Photosensitivity Eruptions Photosensitivity eruptions are usually most marked in sun-exposed areas, but they may extend to sun-protected areas. The mechanism is almost always phototoxicity. Phototoxic reactions resemble sunburn and can occur with first exposure to a drug. Blistering may occur in drug-related pseudoporphyria, most commonly with NSAIDs. The severity of the reaction depends on the tissue level of the drug, its efficiency as a photosensitizer, and the extent of exposure to the activating wavelengths of ultraviolet (UV) light ([Chap. 61](#)).

Common orally administered photosensitizing drugs include fluoroquinolones, tetracycline antibiotics, and trimethoprim/sulfamethoxazole. Other drugs less frequently implicated are chlorpromazine, thiazides, NSAIDs, and BRAF inhibitors. Voriconazole may result in severe photosensitivity, accelerated photoaging, and cutaneous carcinogenesis.

Because UV-A and visible light, which trigger these reactions, are not easily absorbed by nonopaque sunscreens and are transmitted through window glass, photosensitivity reactions may be difficult to block. Photosensitivity reactions abate with removal of either the drug or UV radiation, use of sunscreens that block UV-A light, and treatment of the reaction as one would a sunburn. Rarely, individuals develop persistent reactivity to light, necessitating long-term avoidance of sun exposure. Some chemotherapeutic agents, such as methotrexate, can induce a UV-recall reaction characterized by an erythematous, slightly scaly eruption at sites of prior severe sun exposure.

Pigmentation Changes Drugs, either systemic or topical, may cause a variety of pigmentary changes in the skin by triggering melanocyte production of melanin (as in the case of oral contraceptives causing melasma) or due to deposition of drug or drug metabolites. Long-term minocycline and amiodarone may cause blue-gray pigmentation. Phenothiazine, gold, and bismuth result in gray-brown pigmentation of sun-exposed areas. Numerous cancer chemotherapeutic agents



FIGURE 60-1 Warfarin necrosis involving the breasts.

may be associated with characteristic patterns of pigmentation (e.g., bleomycin, busulfan, daunorubicin, cyclophosphamide, hydroxyurea, fluorouracil, and methotrexate). Clofazimine causes a drug-induced lipofuscinosis with characteristic red-brown coloration. Hyperpigmentation of the face, mucous membranes, and pretibial and subungual areas occurs with antimalarials. Quinacrine causes generalized yellow discoloration. Pigmentation changes may also occur in mucous membranes (busulfan, bismuth), conjunctiva (chlorpromazine, thioridazine, imipramine, clomipramine), nails (zidovudine, doxorubicin, cyclophosphamide, bleomycin, fluorouracil, hydroxyurea), hair, and teeth (tetracyclines).

Warfarin Necrosis of Skin This rare reaction (0.01–0.1%) usually occurs between the third and tenth days of therapy with warfarin, usually in women. Common sites are breasts, thighs, and buttocks (**Fig. 60-1**). Lesions are sharply demarcated, erythematous, or purpuric, and may progress to form large, hemorrhagic bullae with necrosis and eschar formation.

Warfarin anticoagulation in protein C or S deficiency causes an additional reduction in already low circulating levels of endogenous anticoagulants, permitting hypercoagulability and thrombosis in the cutaneous microvasculature, with consequent areas of necrosis. Heparin-induced necrosis may have clinically similar features but is probably due to heparin-induced platelet aggregation with subsequent occlusion of blood vessels; it can affect areas adjacent to the injection site or more distant sites if infused. Levamisole-tainted cocaine (and more recently, heroin) can induce similar skin necrosis; however, the distribution tends to involve the ears and cheeks predominantly, with stellate or retiform purpura. Patients may have abnormal white blood cell counts and may be dual P- and C-ANCA positive.

Drug-Induced Hair Disorders • DRUG-INDUCED HAIR LOSS Medications may affect hair follicles at two different phases of their growth cycle: anagen (growth) or telogen (resting). *Anagen effluvium* occurs within days of drug administration, especially with antimetabolite or other chemotherapeutic drugs. In contrast, in *telogen effluvium*, the delay is 2–4 months following initiation of a new medication. Both present as diffuse, nonscarring alopecia most often reversible after discontinuation of the responsible agent.

A considerable number of drugs have been associated with hair loss. These include antineoplastic agents (alkylating agents, bleomycin, vinca alkaloids, platinum compounds), anticonvulsants (carbamazepine, valproate), beta blockers, antidepressants, antithyroid drugs, IFNs, oral contraceptives, and cholesterol-lowering agents.

DRUG-INDUCED HAIR GROWTH Medications may also cause hair growth. Hirsutism is an excessive growth of terminal hair with masculine hair growth pattern in a female, most often on the face and trunk, due to androgenic stimulation of hormone-sensitive hair follicles (anabolic steroids, oral contraceptives, testosterone, corticotropin). Hypertrichosis is a distinct pattern of hair growth, not in a masculine pattern, typically located on the forehead and temporal regions of the face. Drugs responsible for hypertrichosis include anti-inflammatory drugs, glucocorticoids, vasodilators (diazoxide, minoxidil), diuretics

(acetazolamide), anticonvulsants (phenytoin), immunosuppressive agents (cyclosporine A), psoralens, and zidovudine.

Changes in hair color or structure are uncommon adverse effects from medications. Hair discoloration may occur with chloroquine, IFN- α , chemotherapeutic agents, and tyrosine kinase inhibitors. Changes in hair structure have been observed in patients given EGFR inhibitors, BRAF inhibitors, tyrosine kinase inhibitors, and acitretin.

Drug-Induced Nail Disorders Drug-related nail disorders usually involve all 20 nails and need months to resolve after withdrawal of the medication. The pathogenesis is most often toxic. Drug-induced nail changes include Beau's line (transverse depression of the nail plate), onycholysis (detachment of the distal part of the nail plate), onychomadesis (detachment of the proximal part of the nail plate), pigmentation, and paronychia (inflammation of periungual skin).

ONYCHOLYSIS Onycholysis occurs with tetracyclines, fluoroquinolones, retinoids, NSAIDs, and others, including many chemotherapeutic agents, and may be triggered by exposure to sunlight.

ONYCHOMADESIS Onychomadesis is caused by temporary arrest of nail matrix mitotic activity. Common drugs reported to induce onychomadesis include carbamazepine, lithium, retinoids, and chemotherapeutic agents such as taxanes.

PARONYCHIA Paronychia and multiple pyogenic granulomas with progressive and painful periungual abscess of fingers and toes are side effects of systemic retinoids, lamivudine, indinavir, and anti-EGFR monoclonal antibodies.

NAIL DISCOLORATION Some drugs—including anthracyclines, taxanes, fluorouracil, psoralens, and zidovudine—may induce nail bed hyperpigmentation through melanocyte stimulation. It appears to be reversible and dose dependent.

Toxic Erythema of Chemotherapy and Other Chemotherapy Reactions Because many agents used in cancer chemotherapy inhibit cell division, rapidly proliferating elements of the skin, including hair, mucous membranes, and appendages, are sensitive to their effects. A broad spectrum of chemotherapy-related skin toxicities has been reported, including neutrophilic eccrine hidradenitis, sterile cellulitis, exfoliative dermatitis, and flexural erythema; recent nomenclature classifies these under the unifying diagnosis of toxic erythema of chemotherapy (TEC) (**Fig. 60-2**). Acral erythema is marked by dysesthesia and an erythematous, edematous eruption of the palms and soles. Common causes include cytarabine, doxorubicin, methotrexate, hydroxyurea, fluorouracil, and capecitabine.

The recent introduction of many new monoclonal antibody and small molecular signaling inhibitors for the treatment of cancer has been accompanied by numerous reports of skin and hair toxicity; only the most common of these are mentioned here. EGFR antagonists induce follicular eruptions and nail toxicity after a mean interval of 10 days in a majority of patients. Xerosis, eczematous eruptions, acneiform eruptions, and pruritus are common. Erlotinib is associated with marked hair textural changes. Sorafenib, a tyrosine kinase inhibitor, may result in follicular eruptions and focal bullous eruptions at palmoplantar, flexural sites or areas of frictional pressure. BRAF inhibitors are associated with photosensitivity, palmoplantar hyperkeratosis, hair curling, dyskeratotic (Grover's-like) rash, hyperkeratotic benign cutaneous neoplasms, and keratoacanthoma-like squamous cell carcinomas. Rash, pruritus, and vitiliginous depigmentation have been reported in association with ipilimumab (anti-CTLA4) treatment. Up to 50% of patients experience immune-mediated skin eruptions, including granulomatous reactions, dermatomyositis, panniculitis, and vasculitis. The checkpoint inhibitor class of drugs (including anti-CTLA4, anti-PD-1, and anti-PD-L1 agents) can induce a wide range of cutaneous eruptions beyond vitiligo, including lichenoid, eczematous, granulomatous, papulosquamous, and panniculitis eruptions.

■ IMMUNE CUTANEOUS REACTIONS: COMMON

Maculopapular Eruptions Morbilliform or maculopapular eruptions (**Fig. 60-3**) are the most common of all drug-induced reactions,



FIGURE 60-2 Toxic erythema of chemotherapy.

often start on the trunk or intertriginous areas, and consist of blanching erythematous macules and papules that are symmetric and confluent. Nonblanching, dusky, or bright-red macules as well as mucosal involvement should raise concern for a more severe reaction. Facial involvement in morbilliform eruptions is also uncommon, and the presence of extensive facial lesions with facial edema suggests DIHS. Diagnosis of morbilliform eruptions is rarely assisted by laboratory testing or skin biopsy.

Morbilliform eruptions may be associated with moderate to severe pruritus and fever. A viral exanthem is another differential diagnostic consideration, especially in children, and graft-versus-host disease is also a consideration in the proper clinical setting. Absence of enanthems; absence of ear, nose, throat, and upper respiratory tract symptoms; and polymorphism of the skin lesions support a drug rather than a viral eruption. Common offenders include aminopenicillins, cephalosporins, antibacterial sulfonamides, allopurinol, and antiepileptic drugs. Beta blockers, calcium channel blockers, and ACE inhibitors are rarely the culprit; however, any drug can cause a morbilliform exanthem. Certain medications carry very high rates of morbilliform eruption, including nevirapine and lamotrigine, even in the absence of DIHS reactions. Lamotrigine morbilliform rash is associated with higher starting doses,



FIGURE 60-3 Morbilliform drug eruption.

rapid dose escalation, concomitant use of valproate (which increases lamotrigine levels and half-life), and use in children.

Maculopapular reactions usually develop within 1 week of initiation of therapy and last less than 2 weeks. Occasionally, these eruptions resolve despite continued use of the responsible drug. Because the eruption may also worsen, the suspect drug should be discontinued unless it is essential. It is important to note that the rash may continue to progress for a few days up to 1 week following medication discontinuation. Oral antihistamines and emollients may help relieve pruritus. Short courses of potent topical glucocorticoids can reduce inflammation and symptoms. Systemic glucocorticoid treatment is rarely indicated.

Pruritus Pruritus is associated with almost all drug eruptions and, in some cases, may represent the only symptom of the adverse cutaneous reaction. It may be alleviated by antihistamines such as hydroxyzine or diphenhydramine. Pruritus stemming from specific medications may require distinct treatment, such as selective opiate antagonists for opiate-related pruritus.

Urticaria/Angioedema/Anaphylaxis

Urticaria, the second most frequent type of cutaneous reaction to drugs, is characterized by pruritic, red wheals of varying size rarely lasting more than 24 hours. It has been observed in association with nearly all drugs, most frequently ACE inhibitors, aspirin, NSAIDs, penicillin, and blood products. However, medications account for no more than 10–20% of acute urticaria cases. Deep edema within dermal and subcutaneous tissues is known as angioedema and may involve respiratory and gastrointestinal mucous membranes. Urticaria and angioedema may be part of a life-threatening anaphylactic reaction.

Drug-induced urticaria may be caused by three mechanisms: an IgE-dependent mechanism, circulating immune complexes (serum sickness), and nonimmunologic activation of effector pathways. IgE-dependent urticarial reactions usually occur within 36 hours of drug exposure but can occur within minutes. Immune complex-induced urticaria associated with serum sickness reactions usually occurs 6–12 days after first exposure. In this syndrome, the urticarial eruption (typically polycyclic plaques over distal joints) may be accompanied by fever, hematuria, arthralgias, hepatic dysfunction, and neurologic symptoms. Certain drugs, such as NSAIDs, ACE inhibitors, angiotensin II antagonists, radiographic dye, and opiates, may induce urticarial reactions, angioedema, and anaphylaxis in the absence of drug-specific antibodies through direct mast-cell degranulation.

Radiocontrast agents are a common cause of urticaria and, in rare cases, can cause anaphylaxis. High-osmolality radiocontrast media are about five times more likely to induce urticaria (1%) or anaphylaxis than are newer low-osmolality media. About one-third of those with mild reactions to previous exposure react on reexposure. Pretreatment with prednisone and diphenhydramine reduces reaction rates.

The treatment of urticaria or angioedema depends on the severity of the reaction. In severe cases with respiratory or cardiovascular compromise, epinephrine and intravenous glucocorticoids are the mainstay of therapy. For patients with urticaria without symptoms of angioedema or anaphylaxis, drug withdrawal and oral antihistamines are usually sufficient. Future drug avoidance is recommended; rechallenge, especially in individuals with severe reactions, should only occur in an intensive care setting.

Anaphylactoid Reactions Vancomycin is associated with red man syndrome, a histamine-related anaphylactoid reaction characterized by flushing, diffuse maculopapular eruption, and hypotension. In rare cases, cardiac arrest may be associated with rapid intravenous (IV) infusion of the medication.



FIGURE 60-4 Allergic contact dermatitis (bulous) due to adhesive tape.

Irritant/Allergic Contact Dermatitis Patients using topical medications may develop an irritant or allergic contact dermatitis to the medication itself or to a preservative or other component of the formulation. Reactions to neomycin sulfate, bacitracin, and polymyxin B are common. Contact dermatitis may be seen to adhesives tapes, leading to irritation or blisters around ports and IV sites (**Fig. 60-4**). Harsh disinfectant skin cleansers may lead to localized irritant dermatitis.

Fixed Drug Eruptions These less common reactions are characterized by one or more sharply demarcated, dull red to brown lesions, sometimes with central dusky violaceous erythema and central bulla (**Fig. 60-5**). Hyperpigmentation often results after resolution of the acute inflammation. With rechallenge, the process recurs in the same (fixed) location but may spread to new areas as well. Lesions often involve the lips, hands, legs, face, genitalia, and oral mucosa, and cause a burning sensation. Most patients have multiple lesions. Fixed drug eruptions have been associated with pseudoephedrine (frequently a nonpigmenting reaction), phenolphthalein (in laxatives), sulfonamides, tetracyclines, NSAIDs, barbiturates, and others.

■ IMMUNE CUTANEOUS REACTIONS: RARE AND SEVERE

Drug-Induced Hypersensitivity Syndrome DIHS is a systemic drug reaction also known as DRESS (drug reaction with eosinophilia and systemic symptoms) syndrome; because eosinophilia is not always present, the term *DIHS* is preferred. Clinically, DIHS presents with a prodrome of fever and flu-like symptoms for several days, followed by



FIGURE 60-5 Fixed drug eruption.



FIGURE 60-6 Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS). (Courtesy of Gildo Micheletti, MD.)

the appearance of a diffuse morbilliform eruption, usually involving the face (**Fig. 60-6**). Facial swelling and hand/foot swelling are often present. Systemic manifestations include lymphadenopathy, fever, and leukocytosis (often with eosinophilia or atypical lymphocytosis), as well as hepatitis, nephritis, pneumonitis, myositis, and gastroenteritis, in descending order. Distinct patterns of timing of onset and organ involvement may exist. For example, allopurinol classically induces DIHS with renal involvement; cardiac and lung involvement are more common with minocycline; gastrointestinal involvement is almost exclusively seen with abacavir; and some medications typically do not induce eosinophilia (abacavir, dapsone, lamotrigine). The cutaneous reaction usually begins 2–8 weeks after the drug is started and persists after drug cessation. Signs and symptoms may continue for several weeks, especially those associated with hepatitis. The eruption recurs with rechallenge, and cross-reactions among aromatic anticonvulsants, including phenytoin, carbamazepine, and phenobarbital, are common. Other drugs causing DIHS include antibacterial sulfonamides and other antibiotics. Hypersensitivity to reactive drug metabolites, hydroxylamine for sulfamethoxazole and arene oxide for aromatic anticonvulsants, may be involved in the pathogenesis of DIHS. Recent research suggests that inciting drugs may reactivate quiescent human herpes viruses, including herpesviruses 6 and 7, EBV, and cytomegalovirus (CMV), resulting in expansion of viral-specific CD8+ T lymphocytes and subsequent end-organ damage. Viral reactivation may be associated with a worse clinical prognosis. Mortality rates as high as 10% have been reported, with most fatalities resulting from liver failure. Systemic glucocorticoids (1.5–2 mg/kg/d prednisone equivalent) should be started and tapered slowly over 8–12 weeks, during which time clinical symptoms and labs (including complete blood count with differential, basic metabolic panel, and liver function tests) should be followed carefully. A steroid-sparing agent such as mycophenolate mofetil, IV immunoglobulin, or cyclosporine may be indicated in cases of rapid recurrence upon steroid taper. In all cases, immediate



FIGURE 60-7 Stevens-Johnson syndrome (SJS).

withdrawal of the suspected culprit drug is required. Given the severe long-term complications of myocarditis, patients should undergo cardiac evaluation in cases of severe DIHS or if heart involvement is suspected due to hypotension or arrhythmia. Patients should be closely monitored for resolution of organ dysfunction and for development of late-onset autoimmune thyroiditis and diabetes (up to 6 months).

Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis
SJS and *TEN* are characterized by blisters and mucosal/epidermal detachment resulting from full-thickness epidermal necrosis in the absence of substantial dermal inflammation. The term *Stevens-Johnson syndrome* (*SJS*) describes cases in which the total body surface area of blistering and eventual detachment is <10% (Fig. 60-7). The term *Stevens-Johnson syndrome/toxic epidermal necrolysis* (*SJS/TEN*) *overlap* is used to describe cases with 10–30% epidermal detachment (Fig. 60-8), and the term *toxic epidermal necrolysis* (*TEN*) is used to describe cases with >30% detachment (Figs. 60-9 and 60-10).

Other blistering eruptions with concomitant mucositis may be confused with *SJS/TEN*. Erythema multiforme (EM) associated with herpes simplex virus is characterized by painful mucosal erosions and



FIGURE 60-9 Toxic epidermal necrolysis, hand.

target lesions, typically with an acral distribution and limited skin detachment. *Mycoplasma* and other respiratory infections in children cause a clinically distinct presentation with prominent mucositis and limited cutaneous involvement. The term *reactive infectious mucocutaneous eruption* (*RIME*) has been proposed to help differentiate this clinical entity, which some believe may be the syndrome originally described by Stevens and Johnson.

Patients with *SJS/TEN* initially present with fever >39°C (102.2°F); sore throat; conjunctivitis; and acute onset of painful dusky, atypical, target-like lesions (Fig. 60-11). Intestinal and upper respiratory tract involvement are associated with a poor prognosis, as are older age and greater extent of epidermal detachment. At least 10% of those with

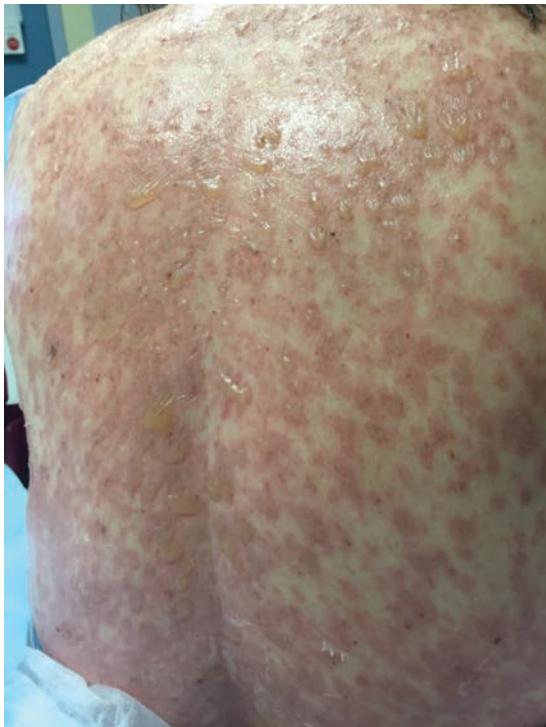


FIGURE 60-8 SJS-TEN overlap.



FIGURE 60-10 Toxic epidermal necrolysis.



FIGURE 60-11 Target-like lesion in SJS.

SJS and 30% of those with TEN die from the disease. Drugs that most commonly cause SJS/TEN are sulfonamides, allopurinol, antiepileptics (e.g., lamotrigine, phenytoin, carbamazepine), oxicam NSAIDs, β -lactam and other antibiotics, and nevirapine. Frozen-section skin biopsy may aid in rapid diagnosis.

At this time, there is no consensus on the most effective treatment for SJS/TEN. The best outcomes stem from early diagnosis, immediate discontinuation of the suspected drug, and meticulous supportive therapy in an intensive care or burn unit. Fluid management, atraumatic wound care, infection prevention and treatment, and ophthalmologic and respiratory support are critical. Early administration of systemic glucocorticoids, intravenous immunoglobulin, cyclosporine, or etanercept may improve disease outcomes, but randomized studies to evaluate potential therapies are lacking and difficult to perform.

Pustular Eruptions AGEP is a rare reaction pattern affecting 3–5 people per million per year. It is thought to be secondary to medication exposure in >90% of cases (Fig. 60-12). Patients typically present with diffuse erythema or erythroderma, as well as high spiking fevers and leukocytosis with neutrophilia. One to two days later, innumerable pinpoint pustules develop overlying the erythema. The pustules are most pronounced in body fold areas; however, they may become generalized and, when coalescent, can lead to superficial erosion. In such cases, differentiating the eruption from SJS in its initial stages may be difficult, although in AGEP, any erosions tend to be more superficial, and prominent mucosal involvement is lacking. Skin biopsy shows collections of neutrophils and sparse necrotic keratinocytes in the upper



FIGURE 60-12 Acute generalized exanthematous pustulosis.

part of the epidermis, unlike the full-thickness epidermal necrosis that characterizes SJS. Before the pustules appear, AGEP may also mimic DIHS due to the prominent fever and erythroderma.

The principal differential diagnosis for AGEP is acute pustular psoriasis, which has an identical clinical and histologic appearance. Many patients with AGEP have a personal or family history of psoriasis. AGEP classically begins within 24–48 hours of drug exposure, although it may occur as much as 1–2 weeks later. β -Lactam antibiotics, calcium channel blockers, macrolide antibiotics, and other inciting agents (including radiocontrast and dialysates) have been reported. Patch testing with the responsible drug often results in a localized pustular eruption.

Overlap Hypersensitivity Syndromes An important concept in the clinical approach to severe drug eruptions is the presence of “overlap syndromes,” most notably DIHS with TEN-like features, DIHS with pustular eruption (AGEP-like), and AGEP with TEN-like features. In several case series of AGEP, 50% of cases had TEN-like or DRESS-like features, and 20% of cases had mucosal involvement resembling SJS/TEN. In one study, up to 20% of all severe drug eruptions had overlap features, suggesting that AGEP, DIHS, and SJS/TEN represent a clinical spectrum with some common pathophysiologic mechanisms. Designation of a single diagnosis based on cutaneous and extracutaneous involvement may not always be possible in cases of hypersensitivity; in such instances, treatment should be geared toward addressing the dominant clinical features. The timing of rash onset with respect to drug administration, which is usually much more delayed in DIHS, and the presence of systemic manifestations such as hepatitis are helpful clues to that diagnosis.

Vasculitis Cutaneous small-vessel vasculitis (CSVV) typically presents with purpuric papules and macules involving the lower extremities and other dependent areas (Fig. 60-13) (Chap. 363). Pustular and hemorrhagic vesicles as well as rounded ulcers also occur. Importantly, vasculitis may involve other organs, including the kidneys,



FIGURE 60-13 Cutaneous small-vessel vasculitis (CSVV, leukocytoclastic vasculitis).

joints, gastrointestinal tract, and lungs, necessitating a thorough clinical evaluation for systemic involvement. Drugs are implicated as a cause of roughly 15% of all cases of small-vessel vasculitis. Antibiotics, particularly β -lactams, are commonly implicated; however, almost any drug can cause vasculitis. Vasculitis may also be idiopathic or due to underlying infection, connective tissue disease, or (rarely) malignancy.

Rare but important types of drug-induced vasculitis include drug-induced ANCA vasculitis. Such patients commonly present with cutaneous manifestations but can develop the full range of symptoms associated with ANCA-associated vasculitis, including crescentic glomerulonephritis and alveolar hemorrhage. Propylthiouracil, methimazole, and hydralazine are common culprits. Drug-induced polyarteritis nodosa has been associated with long-term exposure to minocycline. The presence of perivascular eosinophils on skin biopsy can be a clue to possible drug etiology.

MANAGEMENT OF THE PATIENT WITH SUSPECTED DRUG ERUPTION

There are four main questions to answer regarding a suspected drug eruption:

1. Is the observed rash caused by a medication?
2. Is the reaction severe or evolving with systemic involvement?
3. Which drug or drugs are suspected, and should they be withdrawn?
4. What recommendation can be made for future medication use?

■ EARLY DIAGNOSIS OF SEVERE ERUPTIONS

Rapid recognition of potentially serious or life-threatening reactions is paramount. In this regard, a suspected drug eruption is best defined initially by what it is not (e.g., SJS/TEN, DIHS). **Table 60-2** lists clinical and laboratory features that, if present, suggest the presence of a severe reaction. **Table 60-3** lists the most important of these reactions, along with their key features and commonly associated medications. Any concern for a serious reaction should prompt immediate consultation with a dermatologist and/or referral of the patient to a specialized center.

■ CONFIRMATION OF DRUG REACTION

The probability of drug etiology varies with the pattern of the reaction. Only fixed drug eruptions are always drug-induced. Morbilliform

eruptions are usually viral in children and drug-induced in adults. Among severe reactions, drugs account for 10–20% of anaphylaxis and vasculitis and between 70% and 90% of AGEP, DIHS, SJS, and TEN. Skin biopsy helps characterize the reaction but does not indicate drug causality. Blood counts and liver and renal function tests are important for evaluating organ involvement. The association of mild elevation of liver enzymes and high eosinophil count is frequent but not specific for a drug reaction. Blood tests that could identify an alternative cause, serologic tests (to rule out drug-induced lupus), and serology or polymerase chain reaction for infections may be of great importance to determine a cause.

■ WHAT DRUG(S) TO SUSPECT AND WITHDRAW

Most cases of drug eruptions occur during the first course of treatment with a new medication. A notable exception is IgE-mediated urticaria and anaphylaxis that need presensitization and develop a few minutes to a few hours after rechallenge. Characteristic timing of onset following drug administration is as follows: 4–14 days for morbilliform eruption, 2–4 days for AGEP, 5–28 days for SJS/TEN, and 14–48 days for DIHS. A drug chart, compiling information of all current and past medications/supplements and the timing of administration relative to the rash, is a key diagnostic tool for identifying the inciting drug. Medications introduced for the first time in the relevant time frame are prime suspects. Two other important elements to suspect causality at this stage are (1) previous experience with the drug (or related members of the same pharmacologic class) and (2) alternative etiologic candidates.

The decision to continue or discontinue any medication depends on the severity of the reaction, the severity of the primary disease undergoing treatment, the degree of suspicion of causality, and the feasibility of finding an alternative safer treatment. In any potentially fatal drug reaction, elimination of all possible suspect drugs or unnecessary medications should be immediately attempted. Some rashes may resolve when “treating through” a benign drug-related eruption. The decision to treat through an eruption should, however, remain the exception and withdrawal of every suspect drug the general rule. On the other hand, drugs that are not suspected and are important for the patient (e.g., antihypertensive agents) generally should not be quickly withdrawn. This approach may permit judicious use of these agents in the future.

■ RECOMMENDATION FOR FUTURE USE OF DRUGS

The aims are to (1) prevent the recurrence of the drug eruption and (2) avoid compromising future treatment by inaccurately excluding otherwise useful medications.

A thorough assessment of drug causality is based on timing of the reaction, evaluation of other possible causes, and effect of drug withdrawal or continuation. The RegiSCAR group has proposed the Algorithm of Drug Causality for Epidermal Necrolysis (ALDEN) to rank likelihood of drug causality in SJS/TEN; validation of this and other instruments, such as the Naranjo adverse drug reaction probability scale, is limited. Medication(s) with a “definite” or “probable” causality should be contraindicated, a warning card or medical alert tag (e.g., wristband) should be given to the patient, and the drugs should be listed in the patient’s medical chart as allergies.

■ CROSS-SENSITIVITY

Because of possible cross-sensitivity among chemically related drugs, many physicians recommend avoidance of not only the medication that induced the reaction but also all drugs of the same pharmacologic class.

There are two types of cross-sensitivity. Reactions that depend on a pharmacologic interaction may occur with all drugs that target the same pathway, whether the drugs are structurally similar or not. This is the case with angioedema caused by NSAIDs and ACE inhibitors. In this situation, the risk of recurrence varies from drug to drug in a particular class; however, avoidance of all drugs in the class is usually recommended. Immune recognition of structurally related drugs is the second mechanism by which cross-sensitivity occurs. A classic example

TABLE 60-2 Clinical and Laboratory Findings Suggestive of Severe Cutaneous Adverse Drug Reaction

Cutaneous

- Generalized erythema
- Facial edema
- Skin pain
- Palpable purpura
- Dusky or target-like lesions
- Skin necrosis
- Blisters or epidermal detachment
- Positive Nikolsky sign
- Mucous membrane erosions
- Swelling of lips or tongue

General

- High fever
- Enlarged lymph nodes
- Arthralgias or arthritis
- Shortness of breath, hoarseness, wheezing, hypotension

Laboratory Results

- Eosinophil count >1000/ μ L
- Lymphocytosis with atypical lymphocytes
- Abnormal liver or kidney function tests

Source: From JC Roujeau, RS Stern: Severe adverse cutaneous reactions to drugs. *N Engl J Med* 331:1272, 1994. Copyright © 1994 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

TABLE 60-3 Clinical Features of Severe Cutaneous Drug Reactions

DIAGNOSIS	MUCOSAL LESIONS	TYPICAL SKIN LESIONS	FREQUENT SIGNS AND SYMPTOMS	MOST COMMON CULPRIT DRUGS
Stevens-Johnson syndrome (SJS)	Erosions usually at two or more sites	Small blisters form from dusky macules or atypical targets; rare areas of confluence; detachment ≤10% body surface area	Most cases involve fever	Sulfonamides, anticonvulsants, allopurinol, nonsteroidal anti-inflammatory drugs (NSAIDs)
Toxic epidermal necrolysis (TEN) ^a	Erosions usually at two or more sites	Individual lesions like those seen in SJS; confluent dusky erythema; large sheets of necrotic epidermis; total detachment of >30% body surface area	Nearly all cases involve fever, "acute skin failure," leukopenia	Same as for SJS
Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS)	Mucositis reported in as many as 30%	Diffuse, deep red morbilliform eruption with facial involvement; facial and acral swelling	Fever, lymphadenopathy, hepatitis, nephritis, myocarditis, eosinophilia, atypical lymphocytosis	Anticonvulsants, sulfonamides, allopurinol, minocycline
Acute generalized exanthematous pustulosis (AGEP)	Oral erosions in perhaps 20%	Innumerable pinpoint pustules overlying a diffuse erythematous eruption; may develop superficial erosions	High fever, leukocytosis (neutrophilia), hypocalcemia	β-Lactam antibiotics, calcium channel blockers, macrolide antibiotics
Serum sickness or serum sickness-like reaction	Absent	Urticarial serpiginous or polycyclic rash; purpuric eruption along the sides of the feet and hands is characteristic	Fever, arthralgias	Antithymocyte globulin, cephalosporins, monoclonal antibodies
Anticoagulant-induced necrosis	Infrequent	Purpura and necrosis, especially of central, fatty areas	Pain in affected areas	Warfarin, heparin
Angioedema	Often involved	Urticaria or swelling of the central face, other areas	Respiratory distress, cardiovascular collapse	Angiotensin-converting enzyme (ACE) inhibitors, NSAIDs, contrast dye

^aOverlap of SJS and TEN have features of both, and attachment of 10–30% of body surface area may occur.

Source: From JC Roujeau, RS Stern: Severe adverse cutaneous reactions to drugs. N Engl J Med 331:1272, 1994. Copyright © 1994 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

is hypersensitivity to aromatic antiepileptics (barbiturates, phenytoin, carbamazepine) with up to 50% reaction to a second drug in patients who reacted to one. For other drugs, *in vitro* and *in vivo* data have suggested that cross-reactivity exists only between compounds with very similar chemical structures. Sulfamethoxazole-specific lymphocytes may be activated by other antibacterial sulfonamides but not diuretics, antidiabetic drugs, or anti-COX2 NSAIDs with a sulfonamide group. Though it has been previously reported that 10% of patients with penicillin allergies will also develop allergic reactions to cephalosporin class antibiotics, the cross-reactivity is likely much lower, as is the incidence of true penicillin allergy itself, and severe reactions are very rare.

Recent data suggest that although the risk of developing a drug eruption to another drug is increased in persons with a prior reaction, "cross-sensitivity" is probably not the explanation. As an example, those with a history of an allergic-like reaction to penicillin are at greater risk of developing a reaction to antibacterial sulfonamides than to cephalosporins.

These data suggest that the list of drugs to avoid after a drug reaction should be limited to the causative one(s) and to a few very similar medications.

Because of growing evidence that some severe cutaneous reactions to drugs are associated with HLA genes, it is recommended that first-degree family members of patients with severe cutaneous reactions also should avoid causative agents. This may be most relevant for sulfonamides and antiepileptic medications.

ROLE OF TESTING FOR CAUSALITY AND DRUG RECHALLENGE

The usefulness of laboratory tests, skin-prick, or patch testing to determine causality is debated and may be of limited practical value. Many *in vitro* immunologic assays have been developed for research purposes; however, the predictive value of these tests has not been validated in large series of affected patients. In some cases, diagnostic rechallenge may be appropriate, even for drugs with high rates of adverse reactions.

Skin-prick testing has clinical value in specific settings. In patients with a history suggesting immediate IgE-mediated reactions to penicillin, skin-prick testing with penicillins or cephalosporins has proven useful for identifying patients at risk of anaphylactic reactions to these

agents. Negative skin tests do not totally rule out IgE-mediated reactivity; however, the risk of anaphylaxis in response to penicillin administration in patients with negative skin tests is about 1%. In contrast, two-thirds of patients with a positive skin test experience an allergic response upon rechallenge. The skin tests themselves carry a small risk of anaphylaxis.

For patients with delayed-type hypersensitivity, the clinical utility of skin tests remains questionable. At least one of a combination of several tests (prick, patch, and intradermal) is positive in 50–70% of patients with a reaction "definitely" attributed to a single medication. This low sensitivity corresponds to the observation that readministration of drugs with negative skin testing results in eruptions in 17% of cases.

Desensitization can be considered in those with a history of reaction to a medication that must be used again. Efficacy of such procedures has been demonstrated in cases of immediate reaction to penicillin and positive skin tests, anaphylactic reactions to platinum chemotherapy, and delayed reactions to sulfonamides in patients with AIDS. Desensitization is often successful in HIV-infected patients with morbilliform eruptions to sulfonamides but is not recommended in HIV-infected patients who developed erythroderma or a bullous reaction in response to prior sulfonamide exposure. Various protocols are available, including oral and parenteral approaches. Oral desensitization appears to have a lower risk of serious anaphylactic reaction. Desensitization carries the risk of anaphylaxis regardless of how it is performed and should be performed in monitored clinical settings such as an intensive care unit. After desensitization, many patients experience non-life-threatening reactions during therapy with the culprit drug.

REPORTING

Any severe reaction to drugs should be reported to a regulatory agency or to pharmaceutical companies. Because severe reactions are too rare to be detected in premarketing clinical trials, spontaneous reports are of critical importance for early detection of unexpected life-threatening events. To be useful, the report should contain enough details to permit ascertainment of severity and drug causality.

ACKNOWLEDGMENTS

We acknowledge the contribution of Drs. Jean-Claude Roujeau and Robert S. Stern to this chapter in previous editions.

FURTHER READING

- ALFIREVIC A et al: Genetic testing for prevention of severe drug-induced skin rash. *Cochrane Database Syst Rev* 7:CD010891, 2019.
- CORNEJO-GARCIA JA et al: The genetics of drug hypersensitivity reactions. *J Investig Allergol Clin Immunol* 26:222, 2016.
- DUONG TA et al: Severe cutaneous adverse reactions to drugs. *Lancet* 390:196, 2017.
- KO TM et al: Use of HLA-B*5801 genotyping to prevent allopurinol induced severe cutaneous adverse reactions in Taiwan: National prospective cohort study. *BMJ* 351:h4848, 2015.
- LEE S et al: Association of dipeptidyl peptidase 4 inhibitor use with risk of bullous pemphigoid in patients with diabetes. *JAMA Dermatol* 155:172, 2018.
- MAYORGА C et al: In vitro tests for drug hypersensitivity reactions: An ENDA/EAACI Drug Allergy Interest Group position paper. *Allergy* 71:1103, 2016.
- OUESSALAH A et al: Genetic variants associated with drug-induced immediate hypersensitivity reactions: A PRISMA-compliant systematic review. *Allergy* 71:443, 2016.
- PETER JG et al: Severe delayed cutaneous and systemic reactions to drugs: A global perspective on the science and art of current practice. *J Allergy Clin Immunol Pract* 5:547, 2017.
- PETRELLI F et al: Antibiotic prophylaxis for skin toxicity induced by antiepidermal growth factor receptor agents: A systematic review and meta-analysis. *Br J Dermatol* 175:1166, 2016.
- SASSOLAS B et al: ALDEN, an algorithm for assessment of drug causality in Stevens-Johnson syndrome and toxic epidermal necrolysis: Comparison with case-control analysis. *Clin Pharmacol Ther* 88:60, 2010.
- SEMINARIO-VIDAL L et al: Society of Dermatology Hospitalists supportive care guidelines for the management of Stevens-Johnson syndrome/toxic epidermal necrolysis in adults. *J Am Acad Dermatol* 82:1553, 2020.
- SIMONSEN A et al: Cutaneous adverse reactions to anti-PD-1 treatment: A systematic review. *J Am Acad Dermatol* 83:1415, 2020.
- ZIMMERMANN S et al: Systemic immunomodulating therapies for Stevens-Johnson syndrome and toxic epidermal necrolysis: A systematic review and meta-analysis. *JAMA Dermatol* 153:514, 2017.

atmosphere has led to international agreements to reduce production of those chemicals.

Measurements of solar flux showed a 20-fold regional variation in the amount of energy at 300 nm that reaches the earth's surface. This variability relates to seasonal effects, the path that sunlight traverses through ozone and air, the altitude (a 4% increase for each 300 m of elevation), the latitude (increasing intensity with decreasing latitude), and the amount of cloud cover, fog, and pollution.

The major components of the photobiologic action spectrum that can affect human skin include the UV and visible wavelengths between 290 and 700 nm. In addition, the wavelengths beyond 700 nm in the infrared spectrum primarily emit heat and in certain circumstances may exacerbate the pathologic effects of energy in the UV and visible spectra.

The UV spectrum reaching the Earth represents <10% of total incident solar energy and is arbitrarily divided into two major segments, UV-B and UV-A, which constitute the wavelengths from 290 to 400 nm. UV-B consists of wavelengths between 290 and 320 nm. This portion of the photobiologic action spectrum is the most efficient in producing redness or erythema in human skin and thus is sometimes known as the "sunburn spectrum." UV-A includes wavelengths between 320 and 400 nm and is ~1000-fold less efficient in producing skin redness than is UV-B.

The wavelengths between 400 and 700 nm are visible to the human eye. The photon energy in the visible spectrum is not capable of damaging human skin in the absence of a photosensitizing chemical. Without the absorption of energy by a molecule, there can be no photosensitivity. Thus, the *absorption spectrum* of a molecule is defined as the range of wavelengths it absorbs, whereas the *action spectrum* for an effect of incident radiation is defined as the range of wavelengths that evoke the response.

Photosensitivity occurs when a photon-absorbing chemical (*chromophore*) present in the skin absorbs incident energy, becomes excited, and transfers the absorbed energy to various structures or to molecular oxygen.

UV RADIATION (UVR) AND SKIN STRUCTURE AND FUNCTION

Human skin consists of two major compartments: the outer epidermis, which is a stratified squamous epithelium, and the underlying dermis, which is rich in matrix proteins such as collagens and elastin. Both compartments are susceptible to damage from sun exposure. The epidermis and the dermis contain several chromophores capable of absorbing incident solar energy, including nucleic acids, proteins, and lipids. The outermost epidermal layer, the stratum corneum, is a major absorber of UV-B, and <10% of incident UV-B wavelengths penetrate through the epidermis to the dermis. Approximately 3% of radiation below 300 nm, 20% of radiation below 360 nm, and 33% of short visible radiation reach the basal cell layer in untanned human skin. UV-A readily penetrates to the dermis and is capable of altering structural and matrix proteins that contribute to photoaging of chronically sun-exposed skin, particularly in individuals of light complexion. Thus, longer wavelengths can penetrate more deeply into the skin.

Molecular Targets for UVR-Induced Skin Effects Epidermal DNA—predominantly in keratinocytes and in Langerhans cells (dendritic antigen-presenting cells)—absorbs UV-B and undergoes structural changes between adjacent pyrimidine bases (thymine or cytosine), including the formation of cyclobutane dimers and 6,4-photoproducts. These structural changes are potentially mutagenic and are found in nonmelanoma skin cancers (NMSCs), including basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and Merkel cell carcinoma (MCC). They can be repaired by cellular mechanisms that result in their recognition and excision and the restoration of normal base sequences. The efficient repair of these structural aberrations is crucial, since individuals with defective DNA repair are at high risk for the development of cutaneous cancer. For example, patients with xeroderma pigmentosum, an autosomal recessive disorder, have a variably deficient repair of UV-induced photoproducts. The skin of

61

Photosensitivity and Other Reactions to Sunlight

Alexander G. Marneros, David R. Bickers



SOLAR RADIATION

Sunlight is the most visible and obvious source of comfort in the environment. The sun provides the beneficial effects of warmth and vitamin D synthesis. However, acute and chronic sun exposure also has pathologic consequences. Cutaneous exposure to sunlight is a major cause of human skin cancer and can have immunosuppressive effects as well.

The sun's energy reaching the Earth's surface is limited to components of the ultraviolet (UV) spectrum, the visible spectrum, and portions of the infrared spectrum. The cutoff at the short end of the UV spectrum at ~290 nm is due primarily to stratospheric ozone—formed by highly energetic ionizing radiation—that prevents penetration to the earth's surface of the shorter, more energetic, potentially more harmful wavelengths of solar radiation. Indeed, concern about destruction of the ozone layer by chlorofluorocarbons released into the

these patients often shows the dry, leathery appearance of prematurely photoaged skin, and these patients have an increased frequency of skin cancer already in the first two decades of life. Studies in transgenic mice have verified the importance of functional genes that regulate these repair pathways in preventing the development of UV-induced skin cancer. DNA damage to Langerhans cells may also contribute to the known immunosuppressive effects of UV-B (see “Photoimmunology,” later).

In addition to DNA, molecular oxygen is a target for incident solar UVR, leading to the generation of reactive oxygen species (ROS). These ROS can damage skin components through oxidative damage to DNA, oxidation of polyunsaturated fatty acids in lipids (lipid peroxidation), or oxidation of amino acids in proteins, or they can lead to oxidative deactivation of specific enzymes. UVR can also promote increased cross-linking and degradation of dermal matrix proteins and accumulation of abnormal dermal elastin, leading to photoaging changes known as *solar elastosis*.

Cutaneous Optics and Chromophores *Chromophores* are endogenous or exogenous chemicals that can absorb physical energy. Endogenous chromophores are of two types: (1) normal components of skin, including nucleic acids, proteins, lipids, and 7-dehydrocholesterol (the precursor of vitamin D); and (2) components that are synthesized elsewhere in the body and that circulate in the bloodstream and diffuse into the skin, such as porphyrins. Normally, only trace amounts of porphyrins are present in the skin, but, in selected diseases known as the *porphyrias* (Chap. 416), porphyrins are released into the circulation in increased amounts from the bone marrow and/or the liver and are transported to the skin, where they absorb incident energy both in the Soret band (~400 nm; short visible) and, to a lesser extent, in the red portion of the visible spectrum (580–660 nm). This energy absorption results in the generation of ROS that can mediate structural damage to the skin, manifested as erythema, edema, urticaria, or blister formation. It is of interest that photoexcited porphyrins are currently used in the treatment of BCCs and SCCs and their precursor lesions, actinic keratoses. Known as *photodynamic therapy* (PDT), this modality generates ROS in the skin, leading to cell death. Topical photosensitizers used in PDT are the porphyrin precursors 5-aminolevulinic acid and methyl aminolevulinate, which are readily converted to porphyrins in the skin. It is believed that PDT targets tumor cells for destruction more selectively than it targets adjacent nonneoplastic cells. The efficacy of such therapy requires appropriate timing of the application of methyl aminolevulinate or 5-aminolevulinic acid to the affected skin followed by exposure to artificial sources of visible light. High-intensity blue light has been used successfully for PDT of thin actinic keratoses. Red light PDT penetrates more deeply into the skin and is more beneficial in the treatment of superficial BCCs.

Acute Effects of Sun Exposure The acute effects of skin exposure to sunlight include sunburn and vitamin D synthesis.

SUNBURN This painful skin condition is an acute inflammatory response of the skin, predominantly to UV-B. Generally, an individual's ability to tolerate sunlight is inversely proportional to that individual's degree of melanin pigmentation. Melanin, a complex polymer of tyrosine derivatives, is synthesized in specialized epidermal dendritic cells known as *melanocytes* and is packaged into *melanosomes* that are transferred via dendritic processes into *keratinocytes*, thereby providing photoprotection (dissipating the vast majority of absorbed UVR in the skin) and simultaneously darkening the skin. Sun-induced melanogenesis is a consequence of increased tyrosinase activity in melanocytes. Central to the suntan response is the melanocortin-1 receptor (*MC1R*), and mutations in this gene contribute to the wide variation in human skin and hair color; individuals with red hair and fair skin typically have low *MC1R* activity. In the skin, there are two main types of melanin: eumelanin (providing brown and black pigmentation associated with high *MC1R* activity) and pheomelanin (providing red pigmentation associated with low *MC1R* activity). Pheomelanin is a cysteine-containing red polymer of benzothiazine units and has much weaker shielding capacity against UVR compared to eumelanin. This

may explain why individuals with a higher proportion of pheomelanin (red hair/fair skin appearance) have an increased risk of melanoma formation. In addition, pheomelanin may also promote melanoma formation through induction of oxidative damage by amplifying UV-A-induced ROS but also through UVR-independent mechanisms.

The human *MC1R* gene encodes a G protein-coupled receptor that binds α-melanocyte-stimulating hormone (α-MSH), which is secreted in the skin mainly by keratinocytes in response to UVR. The UV-induced expression of this hormone is controlled by the tumor suppressor p53, and absence of functional p53 attenuates the tanning response. Activation of the melanocortin receptor leads to increased intracellular cyclic adenosine 5'-monophosphate (cAMP) and protein kinase A activation, resulting in an increased transcription of the microphthalmia-associated transcription factor (MITF), which stimulates melanogenesis. Since the precursor of α-MSH, proopiomelanocortin produced by keratinocytes, is also the precursor of β-endorphins, UVR may result in not only increased pigmentation but also increased β-endorphin production in the skin, an effect that has been hypothesized to promote sun-seeking behaviors and even mediate addiction to tanning.

The Fitzpatrick classification of human skin phototypes is based on the efficiency of the epidermal-melanin unit, which usually can be ascertained by asking an individual two questions: (1) Do you burn after sun exposure? (2) Do you tan after sun exposure? The answers to these questions permit division of the population into six skin types, varying from type I (always burn, never tan) to type VI (never burn, always tan) (Table 61-1).

Sunburn erythema is due to vasodilation of dermal blood vessels. There is a lag time (usually 4–12 h) between skin exposure to sunlight and the development of visible redness. The action spectrum for sunburn erythema includes UV-B and UV-A, although UV-B is much more efficient than UV-A in evoking the response. However, UV-A may contribute to sunburn erythema at midday, when much more UV-A than UV-B is present in the solar spectrum. The erythema that accompanies the inflammatory response induced by UVR results from the orchestrated release of cytokines along with growth factors and the generation of ROS. Furthermore, UV-induced activation of nuclear factor κB-dependent gene transcription can augment release of several proinflammatory cytokines and vasoactive mediators. These cytokines and mediators accumulate locally in sunburned skin, providing chemotactic factors that attract neutrophils, macrophages, and T lymphocytes, which promote the inflammatory response. UVR also stimulates infiltration of inflammatory cells through induced expression of adhesion molecules such as E-selectin and intercellular adhesion molecule 1 on endothelial cells and keratinocytes. UVR has been shown to activate phospholipase A₂, resulting in increases in eicosanoids such as prostaglandin E₂, which is known to be a potent inducer of sunburn erythema. The role of eicosanoids in this reaction has been verified by studies showing that nonsteroidal anti-inflammatory drugs (NSAIDs) can reduce sunburn erythema.

Epidermal changes in sunburn include the induction of “sunburn cells,” which are keratinocytes undergoing p53-dependent apoptosis as a defense, with elimination of cells that harbor UV-B-induced structural DNA damage.

VITAMIN D SYNTHESIS AND PHOTOCHEMISTRY Cutaneous exposure to UV-B causes photolysis of epidermal 7-dehydrocholesterol,

TABLE 61-1 Skin Type and Sunburn Sensitivity (Fitzpatrick Classification)

TYPE	DESCRIPTION
I	Always burn, never tan
II	Always burn, sometimes tan
III	Sometimes burn, sometimes tan
IV	Sometimes burn, always tan
V	Never burn, sometimes tan
VI	Never burn, always tan

converting it to pre-vitamin D₃, which then undergoes temperature-dependent isomerization to form the stable hormone vitamin D₃. This compound diffuses to the dermal vasculature and circulates to the liver and kidney, where it is converted to the dihydroxylated functional hormone 1,25-dihydroxyvitamin D₃. Vitamin D metabolites from the circulation and those produced in the skin itself can augment epidermal differentiation signaling and inhibit keratinocyte proliferation. These effects are exploited therapeutically in psoriasis with the topical application of synthetic vitamin D analogues. In addition, vitamin D is increasingly thought to have beneficial effects in several other inflammatory conditions, and some evidence suggests that—besides its classic physiologic effects on calcium metabolism and bone homeostasis—it is associated with a reduced risk of various internal malignancies. There is controversy regarding the risk-to-benefit ratio of sun exposure for vitamin D homeostasis. At present, it is important to emphasize that no clear-cut evidence suggests that the use of sunscreens substantially diminishes vitamin D levels. Since aging also substantially decreases the ability of human skin to photocatalytically produce vitamin D₃, the widespread use of sunscreens that filter out UV-B has led to concerns that the elderly might be unduly susceptible to vitamin D deficiency. However, the amount of sunlight needed to produce sufficient vitamin D is small and does not justify the risks of skin cancer and other types of photodamage linked to increased sun exposure or tanning behavior. Nutritional supplementation of vitamin D is a preferable strategy for patients with vitamin D deficiency.

Chronic Effects of Sun Exposure: Nonmalignant The clinical features of photoaging (*dermatoheliosis*) consist of wrinkling, blotchiness, and telangiectasia, as well as a roughened, irregular, “weather-beaten” leathery appearance.

UVR is important in the pathogenesis of photoaging in human skin, and ROS are likely involved. The dermis and its connective tissue matrix are major targets for sun-associated chronic damage that manifests as solar elastosis, a massive increase in thickened irregular masses of abnormal-appearing elastic fibers. Collagen fibers are also abnormally clumped in the deeper dermis of sun-damaged skin. The chromophores, the action spectra, and the specific biochemical events orchestrating these changes are only partially understood, although more deeply penetrating UV-A seems to be primarily involved. Chronologically aged sun-protected skin and photoaged skin share important molecular features, including connective tissue damage and elevated levels of matrix metalloproteinases (MMPs). MMPs are enzymes involved in the degradation of the extracellular matrix. UV-A induces expression of some MMPs, including MMP-1 and MMP-3, leading to increased collagen breakdown. In addition, UV-A reduces type I procollagen messenger RNA (mRNA) expression. Thus, chronic UVR alters the structure and function of dermal collagen both by inhibiting its synthesis and enhancing its breakdown. Based on these observations, it is not surprising that high-dose UV-A phototherapy may have beneficial effects in some patients with localized fibrotic diseases of the skin, such as localized scleroderma.

Chronic Effects of Sun Exposure: Malignant One of the major known consequences of chronic excessive skin exposure to sunlight is NMSC, including SCCs, BCCs and MCCs (Chap. 76). A model for skin cancer induction involves three major steps: initiation, promotion, and progression. Exposure of human skin to sunlight results in *initiation*, a step by which structural (mutagenic) changes in DNA evoke an irreversible alteration in the target cell (keratinocyte) that begins the tumorigenic process. Exposure to a tumor initiator such as UV-B is believed to be a necessary but not a sufficient step in the malignant process, since initiated skin cells not exposed to tumor promoters generally do not develop into tumors. The second stage in tumor development is *promotion*, a multistep process by which chronic exposure to sunlight evokes further changes that culminate in the clonal expansion of initiated cells and cause the development of premalignant growths known as *actinic keratoses*, which may progress to form SCCs. As a result of extensive studies, it seems clear that UV-B is a *complete carcinogen*, meaning that it can act as both a tumor initiator and a tumor promoter. The third and final step in the malignant

process is *malignant conversion* of benign precursors into cancers, a process thought to enhance genetic instability.

On a molecular level, skin carcinogenesis results from the accumulation of gene mutations that cause inactivation of tumor suppressors, activation of oncogenes, or reactivation of cellular signaling pathways that normally are expressed only during embryologic epidermal development that drive cell proliferation. Interestingly, a large number of UV-induced oncogenic driver mutations that are present in SCCs can already be found in aged sun-exposed normal skin, leading to a growth advantage and innumerable precancerous clones carrying cancer-causing mutations. These mutations occur particularly often in genes that affect proliferation of epidermal stem cells (e.g., NOTCH receptor genes). The pattern of oncogenic gene mutations in aged sun-exposed skin shows considerable overlap with the mutations identified in SCCs, while there is little overlap with the mutations identified in BCCs or melanomas. For example, ~20% of normal aged sun-exposed skin cells and ~60% of SCCs carry driver mutations in *NOTCH1*. Additionally, the accumulation of mutations in the tumor-suppressor gene *p53* can also promote skin carcinogenesis. Indeed, the majority of both human and murine UV-induced skin cancers have characteristic UVR-induced *p53* mutations (C → T and CC → TT transitions). Studies in mice have shown that sunscreens can substantially reduce the frequency of these signature mutations in *p53* and inhibit the induction of tumors. The comparison of UVR-induced gene mutations between aged sun-exposed normal skin and SCCs supports the hypothesis of a progressive accumulation of additional oncogenic mutations that eventually lead to the transition from precancerous cell clones to SCCs. It has been estimated that SCCs harbor ~10 times more oncogenic driver mutations per cell than cells in aged sun-exposed normal skin. Furthermore, while aged sun-exposed skin and SCCs carry similar UVR-induced mutations in *p53* or NOTCH receptors, oncogenic mutations in other genes (e.g., *CDKN2A*) were mainly found in SCCs and not in aged sun-exposed skin, which are thus likely to play a critical role in malignant progression.

Compared to SCCs, BCCs carry a distinct mutational profile in specific genes. BCCs harbor inactivating mutations particularly in the tumor-suppressor gene *patched* or activating mutations in the oncogene *smoothened*, which result in the constitutive activation of the sonic hedgehog signaling pathway and increased cell proliferation. There is also evidence linking alterations in the Wnt/β-catenin signaling pathway, which is known to be critical for hair follicle development, to skin cancer as well. Thus, interactions between this pathway and the hedgehog signaling pathway appear to be involved in both skin carcinogenesis and embryologic development of the skin and hair follicles.

Clonal analysis in mouse models of BCC revealed that tumor cells arise from stem cells of the interfollicular epidermis and the upper infundibulum of the hair follicle. These BCC-initiating cells are reprogrammed to resemble embryonic hair follicle progenitors, whose tumor-initiating ability depends on activation of the Wnt/β-catenin signaling pathway.

SCC initiation occurs both in the interfollicular epidermis and in the hair follicle bulge stem cell populations. In mouse models, the combination of mutant K-Ras and p53 is sufficient to induce invasive SCCs from these cell populations.

The transcription factor Myc is important for stem cell maintenance in the skin, and oncogenic activation of Myc has been implicated in the development of BCCs and SCCs.

The third NMSC is MCC, which is named after its resemblance to Merkel cells in the skin. The incidence of MCCs has been increasing in recent years for unknown reasons. The age-adjusted global incidence is about 1 in 100,000. Just like SCC and BCC, patients with MCCs are usually fair-skinned males in the sixth to eighth decades of life who are living in geographic regions with greater solar UVR. These tumors occur predominantly on the head and neck in older individuals and on the trunk in younger people. MCCs also have a higher incidence among immunosuppressed patients. MCCs are aggressive and life-threatening, poorly differentiated neuroendocrine carcinomas. Overall survival at 5 years is around 50% for local disease, 35% for nodal disease, and 15% for metastatic disease. While the majority of MCCs present

locally, nodal and metastatic disease can occur simultaneously. The pathogenesis of MCCs is closely connected to the Merkel cell polyoma virus (MCPyV). It is now recognized that MCCs can either be MCPyV positive or MCPyV negative. MCPyV-negative MCCs manifest high levels of classic UV-induced signature mutations (C to T or CC to TT) and inactivation of tumor suppressor genes, which could explain the growth of these viral-negative lesions. MCPyV-positive tumors are thought to grow secondary to viral integration into the host genome and acquisition of a truncating mutation of the large T antigen that results in the production of viral oncoproteins. The growth of MCPyV-positive tumors may be further promoted by UVR-induced local immunosuppression. Both forms of MCC are immunogenic, and metastatic MCC has been treated in some patients successfully with PD-1/PD-L1 immune checkpoint inhibitors.

In summary, NMSC involves mutations and alterations in multiple genes and pathways that occur as a result of their chronic accumulation driven by exposure to environmental factors such as solar UVR.

Epidemiologic studies have linked excessive sun exposure to an increased risk of NMSCs and melanoma of the skin; the evidence is far more direct for NMSCs (BCCs, SCCs, and MCCs) than for melanoma. Approximately 80% of NMSCs develop on sun-exposed body areas, including the face, neck, and hands. Major risk factors include male sex, childhood sun exposures, older age, fair skin, and residence at latitudes relatively close to the equator. Individuals with darker-pigmented skin have a lower risk of skin cancer than do fair-skinned individuals. More than 2 million individuals in the United States develop NMSC annually, and the lifetime risk that a fair-skinned individual will develop such a neoplasm is estimated at ~15%. The incidence of NMSC in the population is increasing at a rate of 2–3% per year, likely due to earlier detection and increased opportunities for outdoor activities.

The relationship of sun exposure to melanoma development is less direct, but strong evidence supports an association. Clear-cut risk factors include a positive family or personal history of melanoma and multiple dysplastic nevi. Melanomas can occur during adolescence; the implication is that the latent period for tumor growth is shorter than that for NMSC. For reasons that are only partially understood, melanomas are among the most rapidly increasing human malignancies (*Chap. 76*). One potential explanation is the widespread use of indoor tanning. It is estimated that 30 million people tan indoors in the United States annually, including >2 million adolescents. Furthermore, epidemiologic studies suggest that life in a sunny climate from birth or early childhood may increase the risk of melanoma development. In general, risk does not correlate with cumulative sun exposure but may be related to the duration and extent of exposure in childhood.

However, in contrast to NMSCs, melanoma frequently develops in non-sun-exposed skin, and oncogenic mutations in melanoma may also not be UVR-signature mutations. These observations suggest that UVR-independent factors may contribute to melanomagenesis, which is consistent with findings in mouse models showing that pheomelanin is less efficient in protecting against melanoma than is eumelanin and may promote melanoma through UVR-independent mechanisms.

Importantly, mutations in BRAF and NRAS that lead to activation of a growth-promoting signaling cascade are frequently found in melanoma (but not in SCCs or BCCs), which has led to the development of specific inhibitors of this pathway for the treatment of BRAF-mutant melanoma. However, a high mutational load in melanoma may not be equated with a more unfavorable prognosis. Tumor-specific missense mutations in melanomas can result in neoantigens that facilitate an immune response to the tumor cell. A major advance in treating melanoma, termed immune checkpoint blockade, targets inhibitors of cytotoxic T effector function. For example, the PD-1/PD-L1 interaction inhibits tumor cell apoptosis, promotes peripheral T effector cell exhaustion, and induces conversion of T effector cells to regulatory T cells. Checkpoint inhibitor treatment (e.g., with antibodies that inhibit PD-1 or PD-L1) disrupts this interaction and has resulted in a durable and potent immune destruction of melanoma cells in a subset of patients, leading to prolonged survival of patients with locally advanced or metastatic melanoma. It has recently been shown that a high mutational load in melanomas correlates with improved

therapeutic outcome to immune checkpoint blockade, consistent with the hypothesis that acquired missense mutations in the tumor cells lead to neoantigens that increase the vulnerability of these melanoma cells to attack by activated T cells.

GLOBAL CONSIDERATIONS The frequency of skin cancer shows strong geographic variation, depending on the skin phototype of the majority of the population in these geographic areas, but also depending on the intensity of UVR. For example, both melanoma and NMSCs are particularly common in Australia.

Photoimmunology Exposure to solar radiation causes both local and systemic immunosuppression and involves both the innate and adaptive immune systems. Local immunosuppression is defined as inhibition of immune responses to antigens applied at the irradiated site, whereas systemic immunosuppression is defined as inhibition of immune responses to antigens applied at remote, unirradiated sites. An example of local immunosuppression is that human skin exposure to modest doses of UV-B can deplete the epidermal antigen-presenting Langerhans cells, thereby reducing the degree of allergic sensitization to topical application of the potent contact allergen dinitrochlorobenzene at the irradiated skin site. An example of the systemic immunosuppressive effects of higher doses of UVR is the diminished immunologic response to antigens introduced either epicutaneously or intracutaneously at sites remote from the irradiated site.

The major chromophores in the upper epidermis that are known to initiate UV-mediated immunosuppression include DNA, *trans*-urocanic acid, and membrane components. The action spectrum for UV-induced immunosuppression closely mimics the absorption spectrum of DNA. UVR-induced cyclobutane pyrimidine dimers in Langerhans cells may inhibit antigen presentation. The absorption spectrum of epidermal urocanic acid closely mimics the action spectrum for UV-B-induced immunosuppression. Urocanic acid is a metabolic product of the essential amino acid histidine and accumulates in the upper epidermis through breakdown of the histidine-rich protein filaggrin due to the absence of its catabolizing enzyme in keratinocytes. Urocanic acid is synthesized as a *trans*-isomer, and UV-induced *trans-cis* isomerization of urocanic acid in the stratum corneum drives immunosuppression. *Cis*-urocanic acid may exert its immunosuppressive effects through a variety of mechanisms, including inhibition of antigen presentation by Langerhans cells.

Various additional immunomodulatory factors and cytokines have been implicated in UVR-induced systemic immunosuppression, including tumor necrosis factor- α , interleukin 4 (IL-4), interleukin 10 (IL-10), and eicosanoids. Keratinocytes can release multiple immunomodulators as a response to UVR-induced cell damage that result in an immunosuppressive environment. Induction of IL-4-producing natural killer T cells and of regulatory T cells and B cells has been linked to cell-mediated and humoral immunosuppression as a consequence of UVR damage to skin. Moreover, UVR-induced formation of damage-associated molecular patterns (DAMPs) from necrotic keratinocytes can lead to a type I interferon innate immune response via activation of Toll-like receptor signaling.

One important consequence of chronic sun exposure and associated immunosuppression is an enhanced risk of skin cancer. In part, UV-B activates regulatory T cells that suppress antitumor immune responses via IL-10 expression, whereas in the absence of high UV-B exposure, epidermal Langerhans cells present tumor-associated antigens and induce protective immunity, thereby inhibiting skin tumorigenesis. UV-induced DNA damage is a major molecular trigger of this immunosuppressive effect.

Perhaps the most graphic demonstration of the role of long-term immunosuppression in enhancing the risk of NMSC comes from studies of organ transplant recipients who require lifelong immunosuppressive/antirejection drug regimens. More than 50% of organ transplant recipients develop BCCs and SCCs, and these skin cancers are the most common types of malignancies arising in these patients. The important contributory role of UVR for the formation of these skin cancers in immunosuppressed individuals is highlighted by the observation that nonwhite transplant recipients develop these skin cancers far less

often than white transplant recipients. Rates of BCC and SCC increase with the duration and degree of immunosuppression. Transplant recipients ideally should be screened prior to organ transplantation, be monitored closely thereafter, and adhere to rigorous photoprotection measures, including the use of sunscreens and protective clothing as well as sun avoidance. Notably, immunosuppressive drugs that target the mTOR pathway, such as sirolimus and everolimus, may reduce the risk of NMSC in organ transplant recipients compared to that associated with the use of calcineurin inhibitors (cyclosporine and tacrolimus). The latter may contribute to NMSC formation not only through their immunosuppressive effects but also through suppression of p53-dependent cancer cell senescence pathways independent of host immunity.

Whereas the immunosuppressive effects of UVR contribute to skin cancer, UVR can also exacerbate autoimmune and inflammatory diseases of the skin, including systemic lupus erythematosus (SLE). It has been proposed that in SLE UVR-induced damage to DNA may promote autoantibody formation.

■ PHOTOSENSITIVITY DISEASES

The diagnosis of photosensitivity requires elicitation of a careful history to define the duration of signs and symptoms, the length of time between exposure to sunlight and the development of subjective symptoms and visible changes in the skin. The age of onset can also be a helpful diagnostic clue. For example, the acute photosensitivity of erythropoietic protoporphyrina (EPP) almost always begins in infancy or early childhood, whereas the chronic photosensitivity of porphyria cutanea tarda (PCT) typically begins in the fourth and fifth decades of life. A patient's history of exposure to topical and systemic drugs and chemicals may provide important diagnostic clues. Many classes of drugs can cause photosensitivity on the basis of either phototoxicity or photoallergy.

Examination of the skin may offer important clues. Anatomic areas that are naturally protected from direct sunlight, such as the hairy scalp, the upper eyelids, the retroauricular areas, and the infranasal and submental regions, may be spared, whereas exposed areas show characteristic features of the pathologic process. These anatomic localization patterns are often helpful, but not infallible, in making the diagnosis. For example, airborne contact sensitizers that are blown onto the skin may produce dermatitis that can be difficult to distinguish from photosensitivity despite the fact that such material may trigger skin reactivity in areas shielded from direct sunlight.

Many dermatologic conditions may be caused or aggravated by sunlight (Table 61-2). The role of light in evoking these responses may be dependent on genetic abnormalities ranging from well-described defects in DNA repair that occur in xeroderma pigmentosum to the inherited abnormalities in heme synthesis that characterize the porphyrias.

Polymorphous Light Eruption The most common type of photosensitivity disease is *polymorphous light eruption* (PMLE). Many affected individuals may never seek medical attention because the condition is often transient, becoming manifest in the spring with initial sun exposure but then subsiding spontaneously with continuing exposure, a phenomenon known as "hardening." The major manifestations of PMLE include (often intensely) pruritic erythematous papules that may coalesce into plaques in a patchy distribution on exposed areas of the trunk and forearms. The face is usually less affected. Whereas the morphologic skin findings remain similar for each patient with subsequent recurrences, significant interindividual variations in skin findings are characteristic (hence the term *polymorphous*).

A skin biopsy and phototest procedures in which skin is exposed to multiple erythema doses of UV-A and UV-B may aid in the diagnosis. The action spectrum for PMLE is usually within these portions of the solar spectrum.

Whereas the treatment of an acute flare of PMLE may require topical or systemic glucocorticoids, approaches to preventing PMLE are important and include the use of high-SPF broad-spectrum sunscreens as well as the induction of "hardening" by the cautious administration

TABLE 61-2 Classification of Photosensitivity Diseases

TYPE	DISEASE
Genetic	Erythropoietic porphyria Erythropoietic protoporphyrina Porphyria cutanea tarda—familial Variegate porphyria Hepatoerythropoietic porphyria Albinism Xeroderma pigmentosum Rothmund-Thomson syndrome Bloom syndrome Cockayne syndrome Kindler syndrome Phenylketonuria
Metabolic	Porphyria cutanea tarda—sporadic Hartnup disease Kwashiorkor Pellagra Carcinoid syndrome
Phototoxic	Drugs Drugs, plants, food
Photoallergic	Solar urticaria Drug photoallergy Persistent light reaction/chronic actinic dermatitis
Neoplastic and degenerative	Photoaging Actinic keratosis Melanoma and nonmelanoma skin cancer
Idiopathic	Polymorphous light eruption Hydroa aestivale Actinic prurigo
Photoaggravated	Lupus erythematosus Systemic Subacute cutaneous Discoid Dermatomyositis Herpes simplex Lichen planus actinicus Acne vulgaris (aestivale)

of artificial UV-B (broad-band or narrow-band) and/or UV-A radiation or the use of psoralen plus UV-A (PUVA) photochemotherapy for ~4 weeks before initial sun exposure. Such prophylactic phototherapy or photochemotherapy at the beginning of spring may prevent the occurrence of PMLE throughout the summer.

Actinic prurigo is a photo-induced pruritic eruption that shares similarities with PMLE and often occurs in the spring; however, it can persist throughout the summer and extend into the winter months.

Phototoxicity and Photoallergy These photosensitivity disorders are related to the topical or systemic administration of drugs and other chemicals that can act as chromophores. Both reactions require the absorption of energy by a drug or chemical with consequent production of an excited-state photosensitizer that can transfer its absorbed energy to a bystander molecule or to molecular oxygen, thereby generating tissue-destructive chemical species, including ROS.

Phototoxicity is a nonimmunologic reaction that can be caused by a broad range of drugs and chemicals, some of which are listed in Table 61-3. The usual clinical manifestations include erythema

TABLE 61-3 Drugs That May Cause a Phototoxic Reaction

DRUG	TOPICAL	SYSTEMIC
Amiodarone		+
Dacarbazine		+
Fluoroquinolones		+
5-Fluorouracil	+	+
Furosemide		+
Nalidixic acid		+
Phenothiazines		+
Psoralens	+	+
Retinoids	+/-	+
Sulfonamides		+
Sulfonylureas		+
Tetracyclines		+
Thiazides		+
Vinblastine		+

resembling a sunburn reaction that quickly desquamates, or “peels,” within several days. In addition, edema, vesicles, and bullae may occur. A common phototoxic reaction that occurs after contact with plant-derived furocoumarins and exposure to UV-A radiation is called *phytophotodermatitis*.

Photoallergy is much less common and is distinct in that it is an immunopathologic process. The excited-state photosensitizer may create highly unstable haptic free radicals that bind covalently to macromolecules to form a functional antigen (photoallergen) capable of evoking a delayed-type hypersensitivity response. Most photoallergic reactions are initiated by UV-A rather than UV-B exposure. Some drugs and chemicals that can produce photoallergy are listed in **Table 61-4**. The clinical manifestations typically differ from those of phototoxicity in that an intensely pruritic eczematous dermatitis tends to predominate and evolves into lichenified, thickened, “leathery” changes in sun-exposed areas. A small subset (perhaps 5–10%) of patients with photoallergy may develop a persistent exquisite hypersensitivity to light even when the offending drug or chemical is identified and eliminated, a condition known as *persistent light reaction*.

An uncommon type of persistent photosensitivity is known as *chronic actinic dermatitis*. The affected patients are typically elderly men with a long history of preexisting allergic contact dermatitis or photosensitivity. Common photoallergens associated with this condition are sunscreen ingredients and plant photoallergens. These individuals are usually exquisitely sensitive to UV-B, UV-A, and visible wavelengths.

Phototoxicity and photoallergy often can be diagnostically confirmed by phototest procedures. In patients with suspected phototoxicity,

determining the minimal erythema dose (MED) while the patient is exposed to a suspected agent and then repeating the MED after discontinuation of the agent may provide a clue to the causative drug or chemical. Photopatch testing can be performed to confirm the diagnosis of photoallergy. In this simple variant of ordinary patch testing, a series of known photoallergens is applied to the skin in duplicate, and one set is irradiated with a suberythema dose of UV-A. The development of eczematous changes at sites exposed to sensitizer and light is a positive result. The characteristic abnormality in patients with persistent light reaction is a diminished threshold to erythema evoked by UV-B. Patients with chronic actinic dermatitis usually manifest a broad spectrum of UV hyperresponsiveness and require meticulous photoprotection, including avoidance of sun exposure, use of high-SPF (>30) sunscreens, and, in severe cases, systemic immunosuppression, such as with azathioprine.

The management of drug photosensitivity involves first and foremost the elimination of exposure to the chemical agents responsible for the reaction and the minimization of sun exposure. The acute symptoms of phototoxicity may be ameliorated by cool moist compresses, topical glucocorticoids, and systemically administered NSAIDs. In severely affected individuals, a tapered course of systemic glucocorticoids may be useful. Judicious use of analgesics may be necessary.

Photoallergic reactions require a similar management approach. Furthermore, patients with persistent light reaction and chronic actinic dermatitis must be meticulously protected against light exposure. In selected patients to whom chronic systemic high-dose glucocorticoids pose unacceptable risks, it may be necessary to employ an immunosuppressive drug such as azathioprine, cyclophosphamide, cyclosporine, or mycophenolate mofetil.

Porphyria The porphyrias (**Chap. 416**) are a group of diseases that have in common inherited or acquired derangements in the synthesis of heme. Heme is an iron-chelated tetrapyrrole or porphyrin, and only the nonmetal chelated porphyrins are potent photosensitizers that absorb light intensely in both the short (400–410 nm) and the long (580–650 nm) portions of the visible spectrum.

Heme cannot be reutilized and must be synthesized continuously. The two body compartments with the largest capacity for its production are the bone marrow and the liver. Accordingly, the porphyrias originate in one or the other of these organs, with an end result of excessive endogenous production of potent photosensitizing porphyrins. The porphyrins circulate in the bloodstream and diffuse into the skin, where they absorb solar energy, become photoexcited, generate ROS, and evoke cutaneous photosensitivity. The mechanism of porphyrin photosensitization is known to be photodynamic, or oxygen-dependent, and is mediated by ROS such as singlet oxygen and superoxide anions.

The group of cutaneous porphyrias can be classified as causing either (1) chronic blistering photosensitivity or (2) acute nonblistering photosensitivity. Chronic cutaneous porphyrias include PCT, congenital erythropoietic porphyria (CEP), hepatoerythropoietic porphyria (HEP), hereditary coproporphyria (HCP), and variegate porphyria (VP). CEP, HEP, and PCT manifest only with cutaneous symptoms, while HCP and VP have acute neurovisceral symptoms in addition to the skin photosensitivity. Acute cutaneous nonblistering porphyrias include EPP and X-linked protoporphyrina (XLP). Representative examples of chronic and acute cutaneous porphyrias are discussed below.

Porphyria cutanea tarda (PCT) is the most common type of porphyria and is associated with decreased activity of the heme pathway enzyme uroporphyrinogen decarboxylase (UROD) to <20% of normal. Increased iron and various acquired factors (e.g., alcohol consumption, estrogens, smoking, hepatitis C or HIV infection) can reduce UROD activity. There are two basic types of PCT: (1) the sporadic or acquired type, generally seen in individuals ingesting ethanol or receiving estrogens; and (2) the inherited type, in which there is autosomal dominant transmission of deficient enzyme activity (resulting in heterozygosity for UROD with a reduction to 50% of UROD enzymatic activity and, thus, predisposing the individual to PCT). Both forms are associated with increased hepatic iron stores.

TABLE 61-4 Drugs That May Cause a Photoallergic Reaction

DRUG	TOPICAL	SYSTEMIC
6-Methylcoumarin	+	
Aminobenzoic acid and esters	+	
Bithionol	+	
Chlorpromazine		+
Diclofenac		+
Fluoroquinolones		+
Halogenated salicylanilides	+	
Hypericin (St. John's wort)	+	+
Musk ambrette	+	
Piroxicam		+
Promethazine		+
Sulfonamides		+
Sulfonylureas		+

In both types of PCT, the predominant feature is chronic photosensitivity characterized by increased fragility of sun-exposed skin, particularly areas subject to repeated trauma such as the dorsa of the hands, the forearms, the face, and the ears. The predominant skin lesions are vesicles and bullae that rupture, producing moist erosions (often with a hemorrhagic base) that heal slowly, with crusting and purplish discoloration of the affected skin. Hypertrichosis, mottled pigmentary change, and scleroderma-like induration are associated features. The diagnosis can be confirmed biochemically by measurement of urinary porphyrin excretion, plasma porphyrin assay, and assay of erythrocyte and/or hepatic UROD. Multiple mutations of the *UROD* gene have been identified in human populations. Some patients with PCT have associated mutations in the *HFE* gene, which is linked to hemochromatosis and leads to increased iron absorption by reducing hepcidin expression; these mutations could contribute to the iron overload precipitating PCT, although iron status as measured by serum ferritin, iron levels, and transferrin saturation is no different from that in PCT patients without *HFE* mutations.

Treatment of PCT consists of repeated phlebotomies to diminish the excessive hepatic iron stores and/or intermittent (twice weekly) low doses of orally administered hydroxychloroquine. This treatment is highly effective for PCT but not suited for treatment of other porphyrias. Long-term remission of the disease can often be achieved if the patient eliminates exposure to porphyrinogenic agents, such as ethanol or estrogens, and avoids sun exposure.

Erythropoietic protoporphyrria (EPP) is an acute nonblistering cutaneous porphyria, originates in the bone marrow, and is due to genetic mutations that in most cases decrease the activity of the mitochondrial enzyme ferrochelatase. The major clinical features include acute photosensitivity characterized by painful burning and stinging of exposed skin that often develops during or just after sun exposure. There may be associated skin swelling and, after repeated episodes, a waxlike scarring.

Detection of increased plasma protoporphyrin (PROTO) helps distinguish EPP from lead poisoning and iron-deficiency anemia, in both of which erythrocyte PROTO levels are elevated in the absence of cutaneous photosensitivity. This can be explained by the fact that metal-chelated PROTO is not a photosensitizer.

Rigorous sunlight protection is essential in the management of EPP. Notably, the U.S. Food and Drug Administration (FDA) has approved a synthetic peptide analogue of α -MSH, afamelanotide, in patients with EPP. This drug increases skin pigmentation through melanogenesis, and patients receiving it tolerate sun exposure without pain for longer periods of time and have an improved quality of life as compared to untreated patients. Interestingly, initial studies suggest that afamelanotide may also be beneficial when combined with narrow-band UV-B in the treatment of patients with vitiligo (in patients with skin phototypes IV–VI). Some studies reported that patients with EPP had a moderate increase in tolerance to sunlight after taking oral β -carotene, which may provide this effect by quenching oxygen free radicals.

An algorithm for managing patients with photosensitivity is presented in Fig. 61-1.

PHOTOPROTECTION

Since photosensitivity of the skin results from exposure to sunlight, it follows that absolute avoidance of sunlight will eliminate these disorders. However, contemporary lifestyles make this approach impractical for most individuals. Thus, better approaches to photoprotection have been sought. Natural photoprotection is provided by structural proteins in the epidermis, particularly keratins and melanin. The amount of melanin and its distribution in cells are genetically regulated, and individuals of darker complexion (skin types IV–VI) are at decreased risk for the development of acute sunburn and cutaneous malignancy. Other forms of photoprotection include clothing and sunscreens. Clothing constructed of tightly woven sun-protective fabrics, irrespective of color, affords substantial protection. Wide-brimmed hats, long sleeves, and trousers all reduce direct exposure.

Sunscreens are now considered over-the-counter drugs, and a monograph from the FDA has recognized category I ingredients as safe

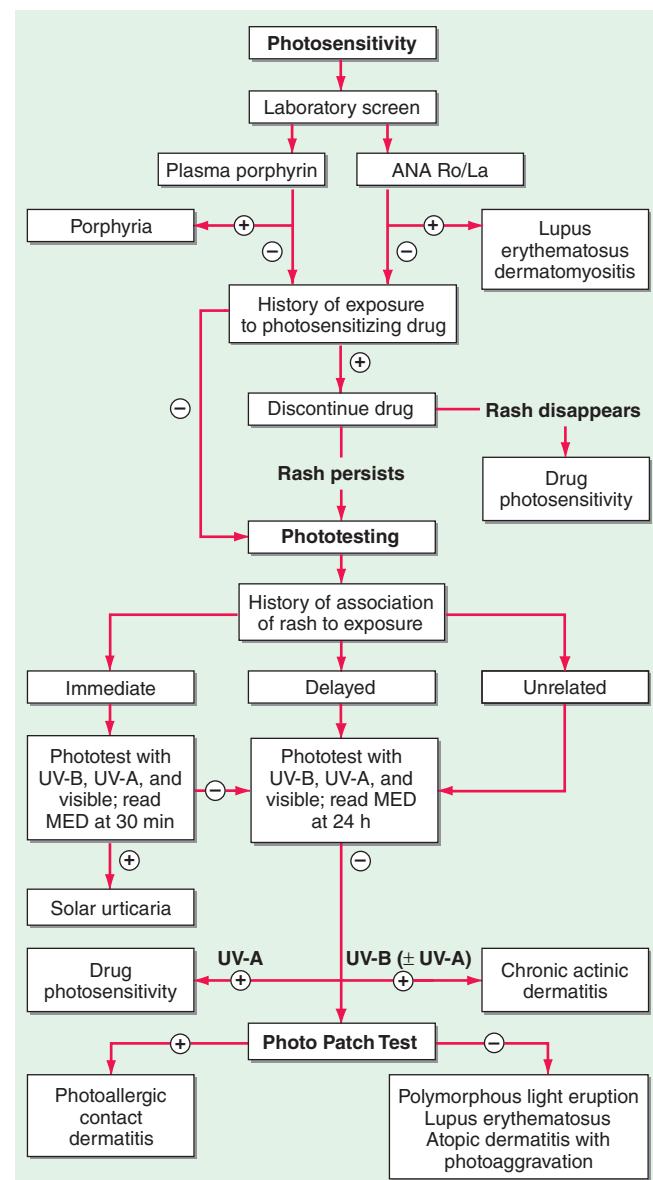


FIGURE 61-1 Algorithm for the diagnosis of a patient with photosensitivity. ANA, antinuclear antibody; MED, minimal erythema dose; UV-A and UV-B, ultraviolet spectrum segments including wavelengths of 320–400 nm and 290–320 nm, respectively.

and effective. Those ingredients are listed in Table 61-5. Sunscreens are rated for their photoprotective effect by their sun protection factor (SPF). The SPF is simply a ratio of the time required to produce sunburn erythema with and without sunscreen application. The SPF of most sunscreens reflects protection from UV-B but not from UV-A. The FDA monograph stipulates that sunscreens must be rated on a scale ranging from minimal (SPF ≥ 2 and < 12) to moderate (SPF ≥ 12 and < 30) to high (SPF ≥ 30 , labeled as 30+).

Broad-spectrum sunscreens contain both UV-B-absorbing and UV-A-absorbing chemicals (organic filters). These chemicals absorb UVR and transfer the absorbed energy to surrounding cells. Among these sunscreen ingredients, cinnamates, PABA derivatives, and salicylates absorb UV-B. Benzophenones or ecamulse (terephthalylidene dicamphor sulfonic acid) offer protection against UV-B and UV-A2, whereas avobenzene protects mainly against UV-A1. In contrast, physical UV blockers (zinc oxide and titanium dioxide) absorb or reflect UVR and offer broad-spectrum protection against UV-B and UV-A. In addition to light absorption, a critical determinant of the sustained photoprotective effect of sunscreens is their water resistance. Sunscreen products with an SPF of 30 or higher, broad-spectrum

TABLE 61-5 FDA Category I Monographed Sunscreen Ingredients

INGREDIENTS	MAXIMUM CONCENTRATION, %
p-Aminobenzoic acid (PABA)	15
Avobenzone	3
Cinoxate	3
Dioxybenzone (benzophenone-8)	3
Ecamsule	15
Homosalate	15
Methyl anthranilate	5
Octocrylene	10
Octyl methoxycinnamate	7.5
Octyl salicylate	5
Oxybenzone (benzophenone-3)	6
Padimate O (octyl dimethyl PABA)	8
Phenylbenzimidazole sulfonic acid	4
Sulisobenzene (benzophenone-4)	10
Titanium dioxide	25
Trolamine salicylate	12
Zinc oxide	25

Abbreviation: FDA, U.S. Food and Drug Administration.

coverage, and water or sweat resistance are recommended for adequate sun protection.

Some degree of photoprotection can be achieved by limiting the time of sun exposure during the day. Since a large part of an individual's total lifetime sun exposure may occur by age 18, it is important to educate parents and young children about the hazards of sunlight. Eliminating exposure at midday will substantially reduce lifetime UVR exposure.

PHOTOTHERAPY AND PHOTOCHEMOTHERAPY

UVR can be used therapeutically. The administration of UV-B alone or in combination with topically applied agents can induce remissions of many dermatologic diseases, including psoriasis, atopic dermatitis, and vitiligo. In particular, narrow-band UV-B treatments (with fluorescent bulbs emitting radiation at ~311 nm) have enhanced efficacy over that obtained with broad-band UV-B in the treatment of psoriasis.

Photochemotherapy in which topically applied or systemically administered psoralens are combined with UV-A (PUVA) is effective in treating psoriasis and the early stages of cutaneous T-cell lymphoma and vitiligo. Psoralens are tricyclic furocoumarins that, when intercalated into DNA and exposed to UV-A, form adducts with pyrimidine bases and eventually form DNA cross-links. These structural changes are thought to decrease DNA synthesis and to be related to the amelioration of psoriasis. Why PUVA photochemotherapy is effective in cutaneous T-cell lymphoma is only partially understood, but it has been shown to induce apoptosis of atypical T lymphocyte populations in the skin. Consequently, direct treatment of circulating atypical lymphocytes by extracorporeal photochemotherapy (photopheresis) has been used in Sézary syndrome as well as in other severe systemic diseases with circulating atypical lymphocytes, such as graft-versus-host disease.

In addition to its effects on DNA, PUVA photochemotherapy stimulates epidermal thickening and melanin synthesis; the latter property, together with its anti-inflammatory effects, provides the rationale for use of PUVA in the depigmenting disease vitiligo. Oral 8-methoxysoralen and UV-A appear to be most effective in this regard, but as many as 100 treatments extending over 12–18 months may be required for satisfactory repigmentation.

Not surprisingly, the major side effects of long-term UV-B phototherapy and PUVA photochemotherapy mimic those seen in individuals with chronic sun exposure. Despite these risks, the therapeutic index of these modalities continues to be excellent. It is important to

choose the most appropriate phototherapeutic approach for a specific dermatologic disease. For example, narrow-band UV-B has been reported in several studies to be as effective as PUVA photochemotherapy in the treatment of psoriasis but to pose a lower risk of skin cancer development than PUVA.

FURTHER READING

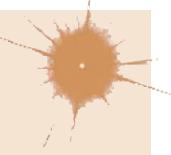
- BERNARD JJ et al: Photoimmunology: How ultraviolet radiation affects the immune system. *Nat Rev Immunol* 11:688, 2019.
- FELL GL et al: Skin beta-endorphin mediates addiction to UV light. *Cell* 157:1527, 2014.
- HARMS PW et al: The biology and treatment of Merkel cell carcinoma: Current understanding and research priorities. *Nat Rev Clin Oncol* 15:763, 2018.
- JANSEN R et al: Photoprotection: Part II. Sunscreen: Development, efficacy, and controversies. *J Am Acad Dermatol* 69:867, 2013.
- LO JA et al: The melanoma revolution: From UV carcinogenesis to a new era in therapeutics. *Science* 346:945, 2014.
- MARTINCORENA I et al: Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348:880, 2015.
- SANCHEZ-DANES A et al: Defining the clonal dynamics leading to mouse skin tumour initiation. *Nature* 536:298, 2016.

Section 9 Hematologic Alterations

62

Interpreting Peripheral Blood Smears

Dan L. Longo



Some of the relevant findings in peripheral blood, enlarged lymph nodes, and bone marrow are illustrated in this chapter. Systematic histologic examination of the bone marrow and lymph nodes is beyond the scope of a general medicine textbook. However, every internist should know how to examine a peripheral blood smear.

The examination of a peripheral blood smear is one of the most informative exercises a physician can perform. Although advances in automated technology have made the examination of a peripheral blood smear by a physician seem less important, the technology is not a completely satisfactory replacement for a blood smear interpretation by a trained medical professional who also knows the patient's clinical history, family history, social history, and physical findings. It is useful to ask the laboratory to generate a Wright's-stained peripheral blood smear and examine it.

The best place to examine blood cell morphology is the feathered edge of the blood smear where red cells lie in a single layer, side by side, just barely touching one another but not overlapping. The author's approach is to look at the smallest cellular elements, the platelets, first and work his way up in size to red cells and then white cells.

Using an oil immersion lens that magnifies the cells 100-fold, one counts the platelets in five to six fields, averages the number per field, and multiplies by 20,000 to get a rough estimate of the platelet count. The platelets are usually 1–2 µm in diameter and have a blue granulated appearance. There is usually 1 platelet for every 20 or so red cells. Of course, the automated counter is much more accurate, but gross disparities between the automated and manual counts should be assessed. Large platelets may be a sign of rapid platelet turnover, as young platelets are often larger than old ones; alternatively, certain rare inherited syndromes can produce large platelets. If the platelet count is low, the absence of large (young) platelets may be an indicator of marrow production problems. Platelet clumping visible on the smear

can be associated with falsely low automated platelet counts. Clumping may be caused by the anticoagulant into which the blood is drawn. Similarly, neutrophil fragmentation can be a source of falsely elevated automated platelet counts. The absence of platelet granules may be an artifact of the handling of the blood or may indicate marrow disease or a rare congenital anomaly, gray platelet syndrome. Elevated platelet counts usually signify a myeloproliferative disorder or a reaction to systemic inflammation.

Next one examines the red blood cells. One can gauge their size by comparing the red cell to the nucleus of a small lymphocyte. Both are normally about 8- μm wide. Red cells that are smaller than the small lymphocyte nucleus may be microcytic; those larger than the small lymphocyte nucleus may be macrocytic. Macrocytic cells also tend to be more oval than spherical in shape and are sometimes called macroovalocytes. The automated mean corpuscular volume (MCV) can assist in making a classification. However, some patients may have both iron and vitamin B₁₂ deficiency, which will produce an MCV in the normal range but wide variation in red cell size. When the red cells vary greatly in size, *anisocytosis* is said to be present. When the red cells vary greatly in shape, *poikilocytosis* is said to be present. The electronic cell counter provides an independent assessment of variability in red cell size. It measures the range of red cell volumes and reports the results as “red cell distribution width” (RDW). This value is calculated from the MCV; thus, cell width is not being measured but cell volume is. The term is derived from the curve displaying the frequency of cells at each volume, also called the distribution. The width of red cell volume distribution curve is what determines the RDW. The RDW is calculated as follows: RDW = (standard deviation of MCV ÷ mean MCV) × 100. In the presence of morphologic anisocytosis, RDW (normally 11–14%) increases to 15–18%. The RDW is useful in at least two clinical settings. In patients with microcytic anemia, the differential diagnosis is generally between iron deficiency and thalassemia. In thalassemia, the small red cells are generally of uniform size with a normal small RDW. In iron deficiency, the size variability and the RDW are large. In addition, a large RDW can suggest a dimorphic anemia when a chronic atrophic gastritis can produce both vitamin B₁₂ malabsorption to produce macrocytic anemia and blood loss to produce iron deficiency. In such settings, RDW is also large. An elevated RDW also has been reported as a risk factor for all-cause mortality in population-based studies, a finding that is unexplained currently.

After red cell size is assessed, one examines the hemoglobin content of the cells. They are either normal in color (*normochromic*) or pale in color (*hypochromic*). They are never “hyperchromic.” If more than the normal amount of hemoglobin is made, the cells get larger—they do not become darker. In addition to hemoglobin content, the red cells are examined for inclusions. Red cell inclusions are the following:

1. *Basophilic stippling*—diffuse fine or coarse blue dots in the red cell usually representing RNA residue—especially common in lead poisoning
2. *Howell-Jolly bodies*—dense blue circular inclusions that represent nuclear remnants—their presence implies defective splenic function
3. *Nuclei*—red cells may be released or pushed out of the marrow prematurely before nuclear extrusion—often implies a myelophthisic process or a vigorous narrow response to anemia, usually hemolytic anemia
4. *Parasites*—red cell parasites include malaria and babesia (Chap. A6)
5. *Polychromatophilia*—the red cell cytoplasm has a bluish hue, reflecting the persistence of ribosomes still actively making hemoglobin in a young red cell

Vital stains are necessary to see precipitated hemoglobin called *Heinz bodies*.

Red cells can take on a variety of different shapes. All abnormally shaped red cells are *poikilocytes*. Small red cells without the central pallor are *spherocytes*; they can be seen in hereditary spherocytosis, hemolytic anemias of other causes, and clostridial sepsis. *Dacrocytes* are teardrop-shaped cells that can be seen in hemolytic anemias, severe iron deficiency, thalassemias, myelofibrosis, and myelodysplastic syndromes. *Schistocytes* are helmet-shaped cells that reflect

microangiopathic hemolytic anemia or fragmentation on an artificial heart valve. *Echinocytes* are spiculated red cells with the spikes evenly spaced; they can represent an artifact of abnormal drying of the blood smear or reflect changes in stored blood. They also can be seen in renal failure and malnutrition and are often reversible. *Acanthocytes* are spiculated red cells with the spikes irregularly distributed. This process tends to be irreversible and reflects underlying renal disease, abetalipoproteinemia, or splenectomy. *Elliptocytes* are elliptical-shaped red cells that can reflect an inherited defect in the red cell membrane, but they also are seen in iron deficiency, myelodysplastic syndromes, megaloblastic anemia, and thalassemias. *Stomatocytes* are red cells in which the area of central pallor takes on the morphology of a slit instead of the usual round shape. Stomatocytes can indicate an inherited red cell membrane defect and also can be seen in alcoholism. *Target cells* have an area of central pallor that contains a dense center, or bull's eye. These cells are seen classically in thalassemia, but they are also present in iron deficiency, cholestatic liver disease, and some hemoglobinopathies. They also can be generated artificially by improper slide making.

One last feature of the red cells to assess before moving to the white blood cells is the distribution of the red cells on the smear. In most individuals, the cells lie side by side in a single layer. Some patients have red cell clumping (called *agglutination*) in which the red cells pile upon one another; it is seen in certain paraproteinemics and autoimmune hemolytic anemias. Another abnormal distribution involves red cells lying in single cell rows on top of one another like stacks of coins. This is called *rouleaux formation* and reflects abnormal serum protein levels.

Finally, one examines the white blood cells. Three types of granulocytes are usually present: neutrophils, eosinophils, and basophils, in decreasing frequency. Neutrophils are generally the most abundant white cell. They are round, are 10–14 μm wide, and contain a lobulated nucleus with two to five lobes connected by a thin chromatin thread. Bands are immature neutrophils that have not completed nuclear condensation and have a U-shaped nucleus. Bands reflect a left shift in neutrophil maturation in an effort to make more cells more rapidly. Neutrophils can provide clues to a variety of conditions. Vacuolated neutrophils may be a sign of bacterial sepsis. The presence of 1- to 2- μm blue cytoplasmic inclusions, called *Döhle bodies*, can reflect infections, burns, or other inflammatory states. If the neutrophil granules are larger than normal and stain a darker blue, “toxic granulations” are said to be present, and they also suggest a systemic inflammation. The presence of neutrophils with more than five nuclear lobes suggests megaloblastic anemia. Large misshapen granules may reflect the inherited Chédiak-Higashi syndrome.

Eosinophils are slightly larger than neutrophils, have bilobed nuclei, and contain large red granules. Diseases of eosinophils are associated with too many of them rather than any morphologic or qualitative change. They normally total less than one-thirtieth the number of neutrophils. Basophils are even more rare than eosinophils in the blood. They have large dark blue granules and may be increased as part of chronic myeloid leukemia.

Lymphocytes can be present in several morphologic forms. Most common in healthy individuals are small lymphocytes with a small dark nucleus and scarce cytoplasm. In the presence of viral infections, more of the lymphocytes are larger, about the size of neutrophils, with abundant cytoplasm and a less condensed nuclear chromatin. These cells are called *reactive lymphocytes*. About 1% of lymphocytes are larger and contain blue granules in a light blue cytoplasm; they are called *large granular lymphocytes*. In chronic lymphoid leukemia, the small lymphocytes are increased in number, and many of them are ruptured in making the blood smear, leaving a smudge of nuclear material without a surrounding cytoplasm or cell membrane; they are called *smudge cells* and are rare in the absence of chronic lymphoid leukemia.

Monocytes are the largest white blood cells, ranging from 15 to 22 μm in diameter. The nucleus can take on a variety of shapes but usually appears to be folded; the cytoplasm is gray.

Abnormal cells may appear in the blood. Most often, the abnormal cells originate from neoplasms of bone marrow-derived cells, including lymphoid cells, myeloid cells, and occasionally red cells. More rarely, other types of tumors can get access to the bloodstream, and rare

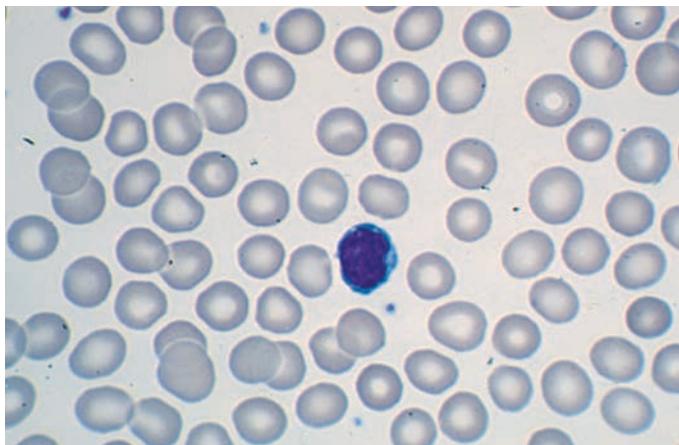


FIGURE 62-1 Normal peripheral blood smear. Small lymphocyte in center of field. Note that the diameter of the red blood cell is similar to the diameter of the small lymphocyte nucleus. (Source: From M Lichtman et al (eds). *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault, *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

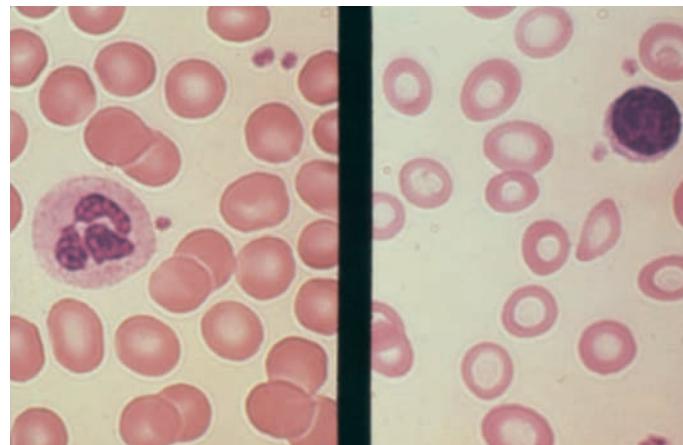


FIGURE 62-4 Iron deficiency anemia next to normal red blood cells. Microcytes (right panel) are smaller than normal red blood cells (cell diameter $<7\text{ }\mu\text{m}$) and may or may not be poorly hemoglobinized (hypochromic). (Source: From M Lichtman et al (eds). *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault, *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

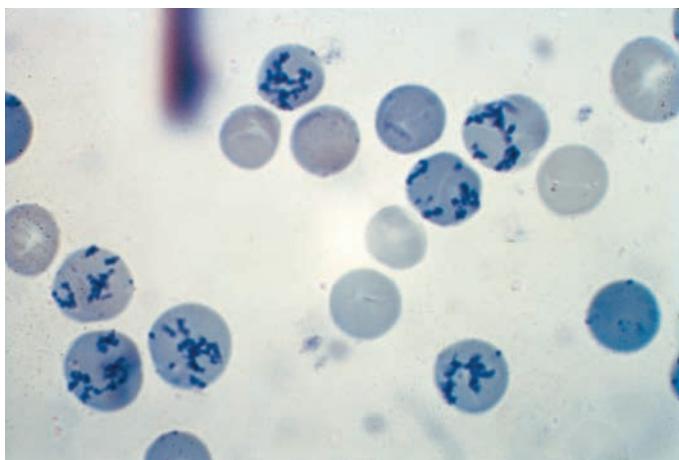


FIGURE 62-2 Reticulocyte count preparation. This new methylene blue-stained blood smear shows large numbers of heavily stained reticulocytes (the cells containing the dark blue-staining RNA precipitates). (Source: From M Lichtman et al (eds). *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

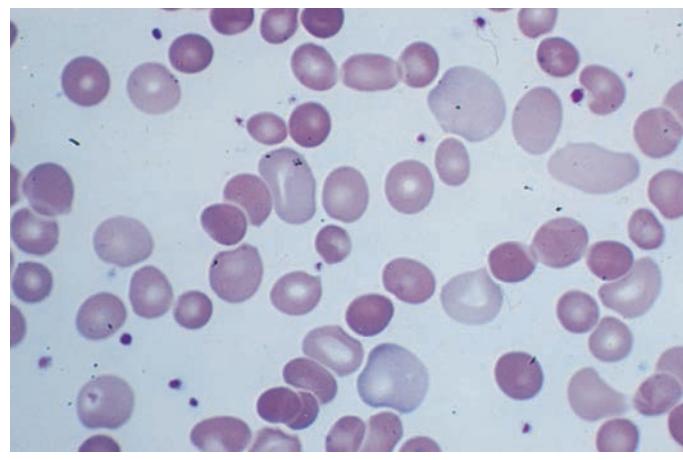


FIGURE 62-5 Polychromatophilia. Note large red cells with light purple coloring. (Source: From M Lichtman et al (eds). *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

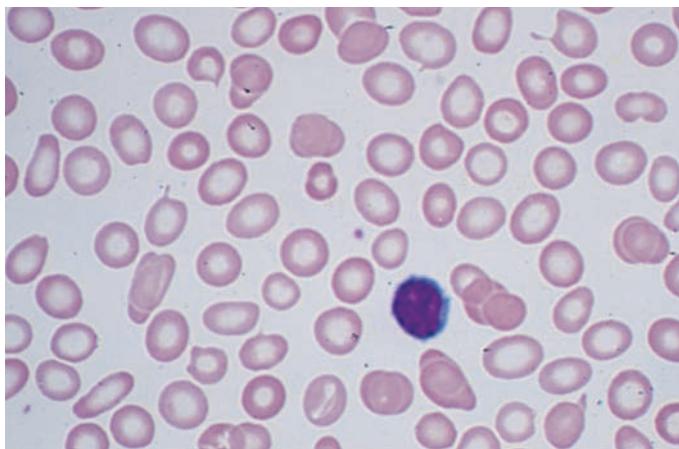


FIGURE 62-3 Hypochromic microcytic anemia of iron deficiency. Small lymphocyte in field helps assess the red blood cell size. (Source: From M Lichtman et al (eds). *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

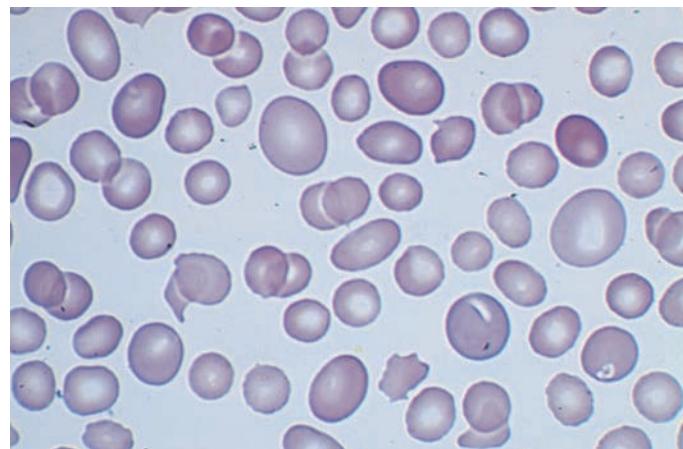


FIGURE 62-6 Macrocytosis. These cells are both larger than normal (mean corpuscular volume >100) and somewhat oval in shape. Some morphologists call these cells macroovalocytes. (Source: From M Lichtman et al (eds). *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

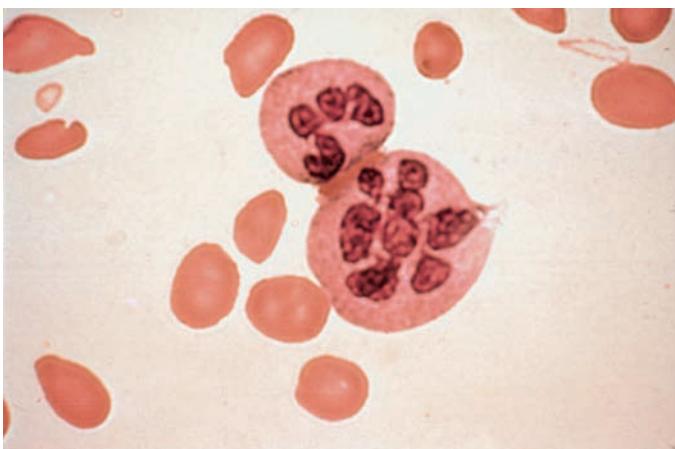


FIGURE 62-7 Hypersegmented neutrophils. Hypersegmented neutrophils (multilobed polymorphonuclear leukocytes) are larger than normal neutrophils with five or more segmented nuclear lobes. They are commonly seen with folic acid or vitamin B₁₂ deficiency. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

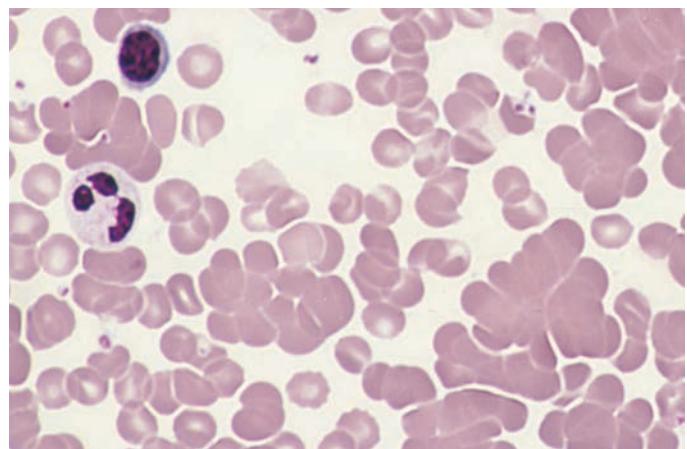


FIGURE 62-10 Red cell agglutination. Small lymphocyte and segmented neutrophil in upper left center. Note irregular collections of aggregated red cells. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)



FIGURE 62-8 Spherocytosis. Note small hyperchromatic cells without the usual clear area in the center. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

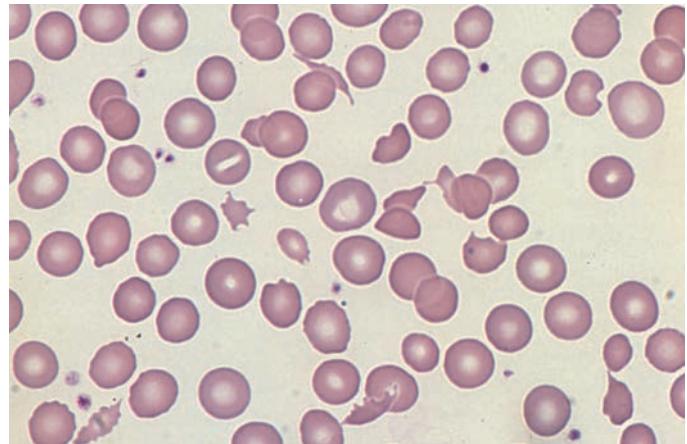


FIGURE 62-11 Fragmented red cells. Heart valve hemolysis. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

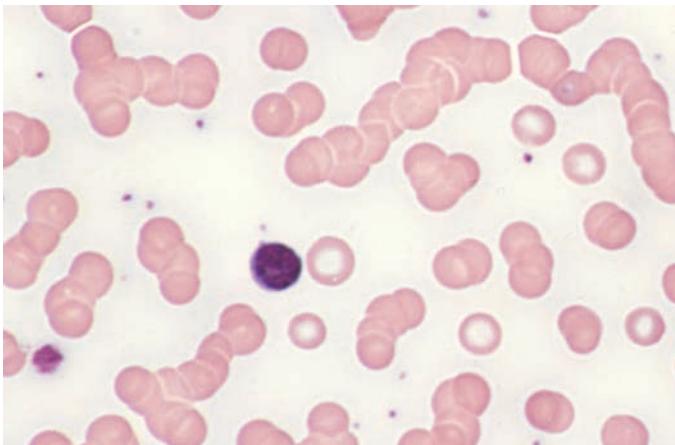


FIGURE 62-9 Rouleaux formation. Small lymphocyte in center of field. These red cells align themselves in stacks and are related to increased serum protein levels. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

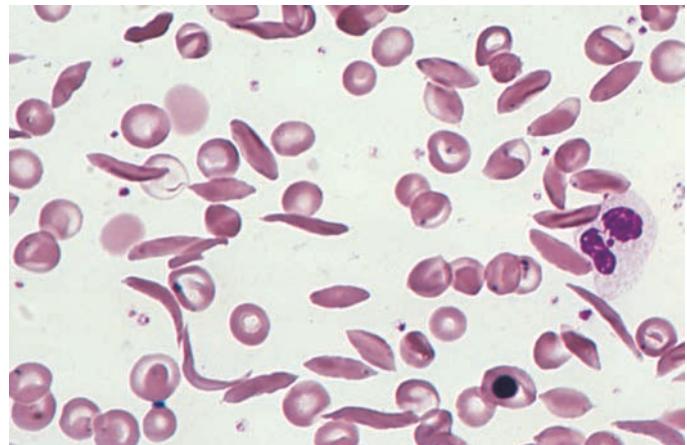


FIGURE 62-12 Sickled cells. Homozygous sickle cell disease. A nucleated red cell and neutrophil are also in the field. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

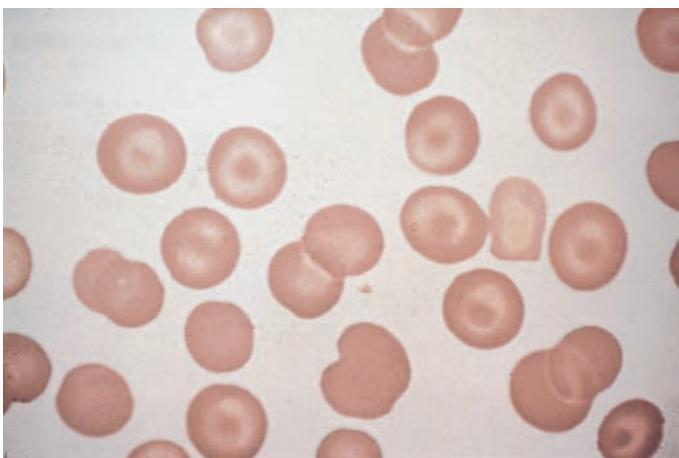


FIGURE 62-13 Target cells. Target cells are recognized by the bull's-eye appearance of the cell. Small numbers of target cells are seen with liver disease and thalassemia. Larger numbers are typical of hemoglobin C disease. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

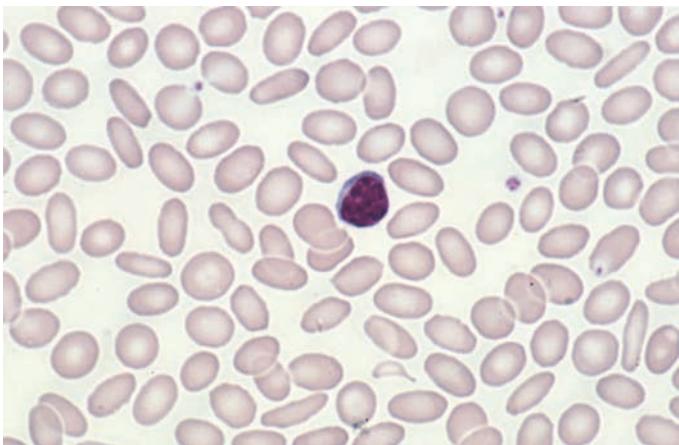


FIGURE 62-14 Elliptocytosis. Small lymphocyte in center of field. Elliptical shape of red cells related to weakened membrane structure, usually due to mutations in spectrin. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

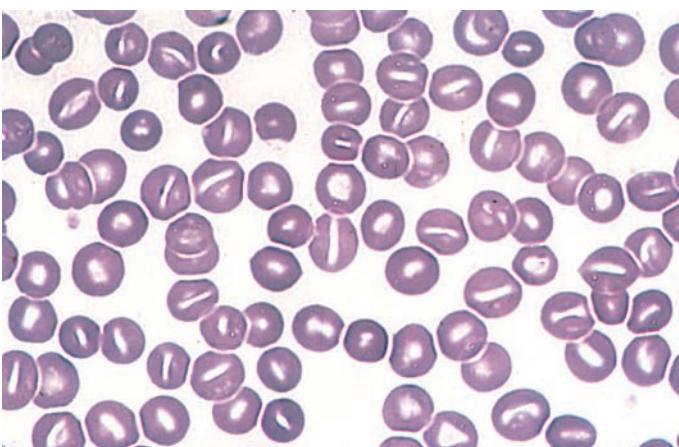


FIGURE 62-15 Stomatocytosis. Red cells characterized by a wide transverse slit or stoma. This often is seen as an artifact in a dehydrated blood smear. These cells can be seen in hemolytic anemias and in conditions in which the red cell is overhydrated or dehydrated. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

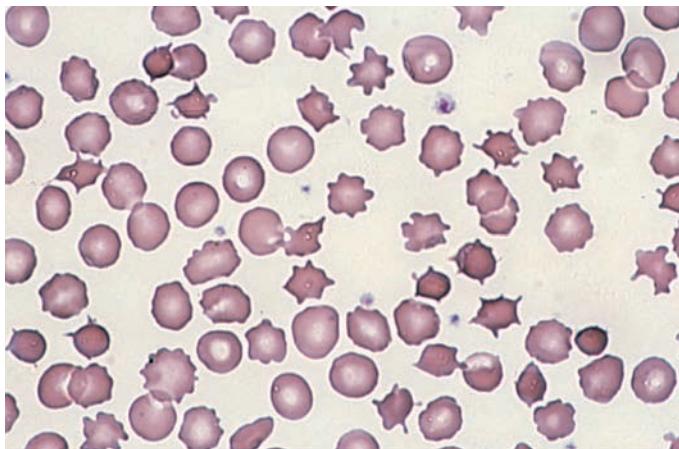


FIGURE 62-16 Acanthocytosis. Spiculated red cells are of two types: *acanthocytes* are contracted dense cells with irregular membrane projections that vary in length and width; *echinocytes* have small, uniform, and evenly spaced membrane projections. Acanthocytes are present in severe liver disease, in patients with abetalipoproteinemia, and in rare patients with McLeod blood group. Echinocytes are found in patients with severe uremia, in glycolytic red cell enzyme defects, and in microangiopathic hemolytic anemia. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

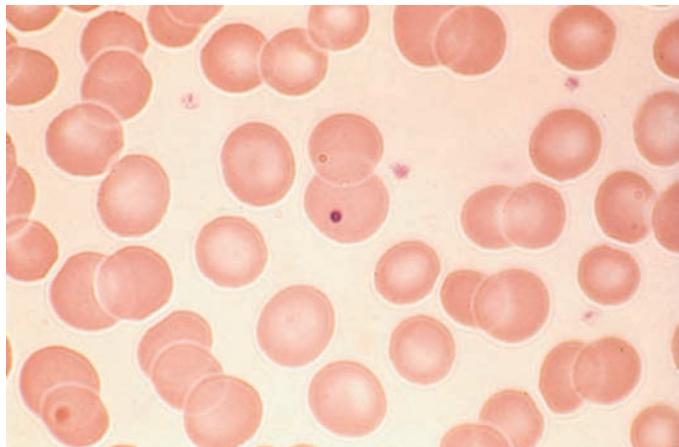


FIGURE 62-17 Howell-Jolly bodies. Howell-Jolly bodies are tiny nuclear remnants that normally are removed by the spleen. They appear in the blood after splenectomy (defect in removal) and with maturation/dysplastic disorders (excess production). (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

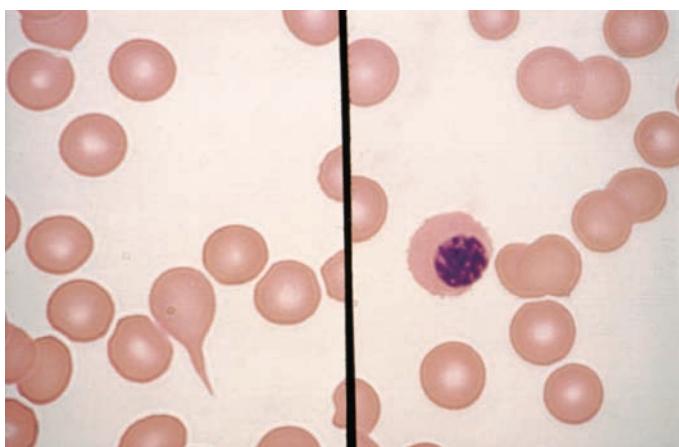


FIGURE 62-18 Teardrop cells and nucleated red blood cells characteristic of myelofibrosis. A teardrop-shaped red blood cell (left panel) and a nucleated red blood cell (right panel) as typically seen with myelofibrosis and extramedullary hematopoiesis. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

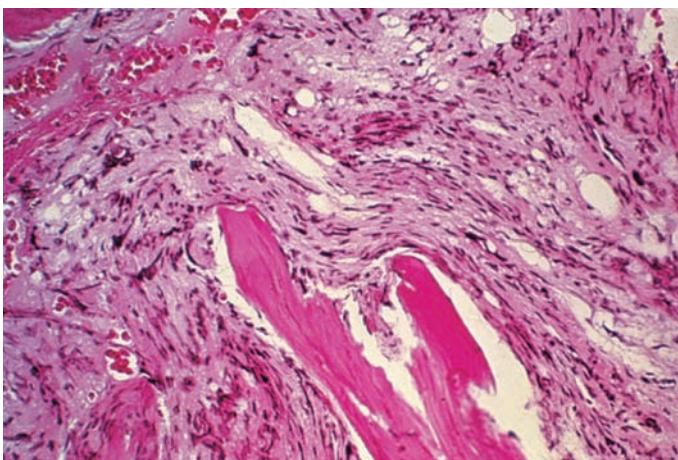


FIGURE 62-19 Myelofibrosis of the bone marrow. Total replacement of marrow precursors and fat cells by a dense infiltrate of reticulin fibers and collagen (hematoxylin and eosin stain). (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

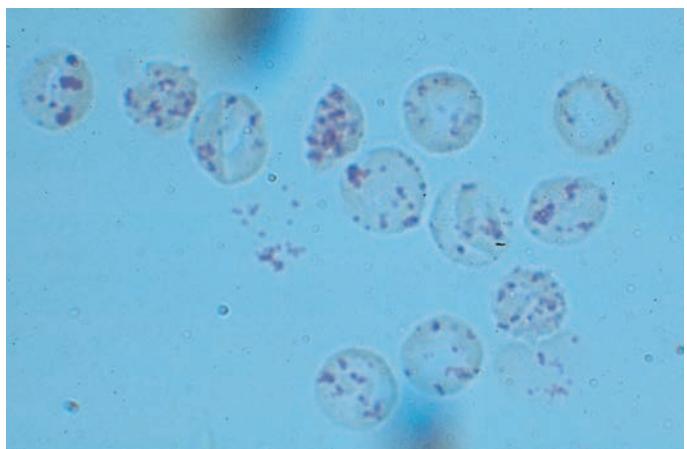


FIGURE 62-22 Heinz bodies. Blood mixed with hypotonic solution of crystal violet. The stained material is precipitates of denatured hemoglobin within cells. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

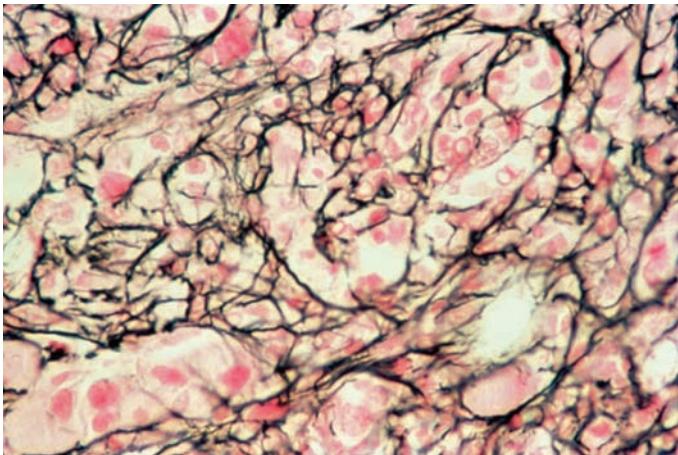


FIGURE 62-20 Reticulin stain of marrow myelofibrosis. Silver stain of a myelofibrotic marrow showing an increase in reticulin fibers (black-staining threads). (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

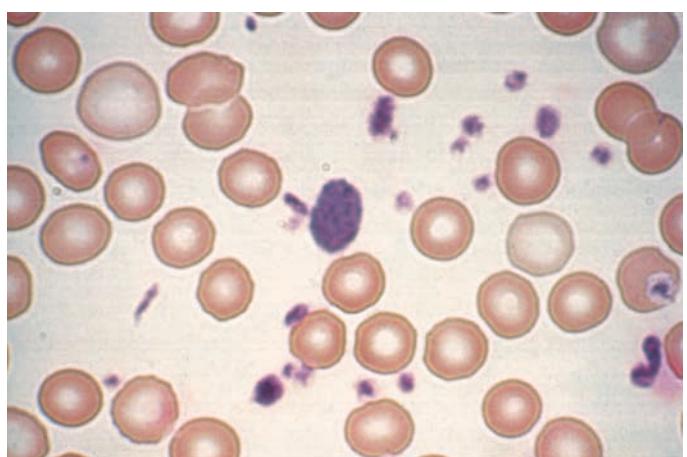


FIGURE 62-23 Giant platelets. Giant platelets, together with a marked increase in the platelet count, are seen in myeloproliferative disorders, especially primary thrombocythemia. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

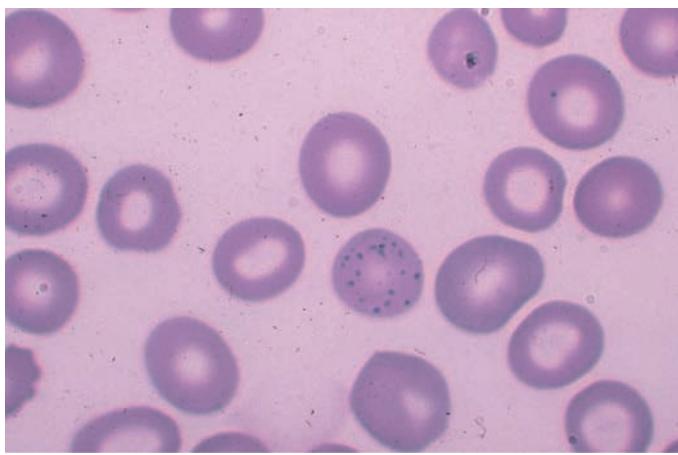


FIGURE 62-21 Stippled red cell in lead poisoning. Mild hypochromia. Coarsely stippled red cell. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

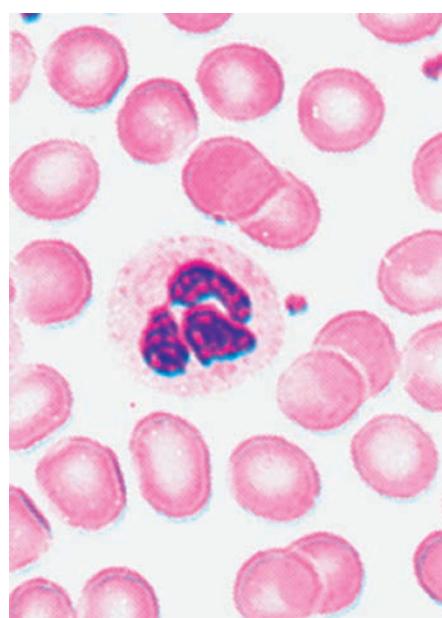


FIGURE 62-24 Normal granulocytes. The normal granulocyte has a segmented nucleus with heavy, clumped chromatin; fine neutrophilic granules are dispersed throughout the cytoplasm. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

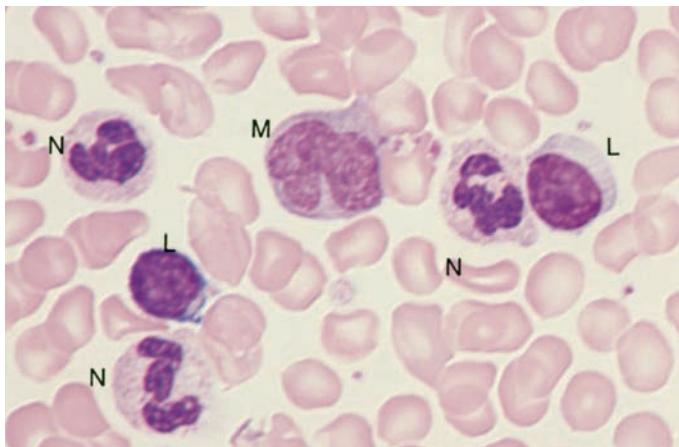


FIGURE 62-25 Normal monocytes. The film was prepared from the buffy coat of the blood from a normal donor. L, lymphocyte; M, monocyte; N, neutrophil. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

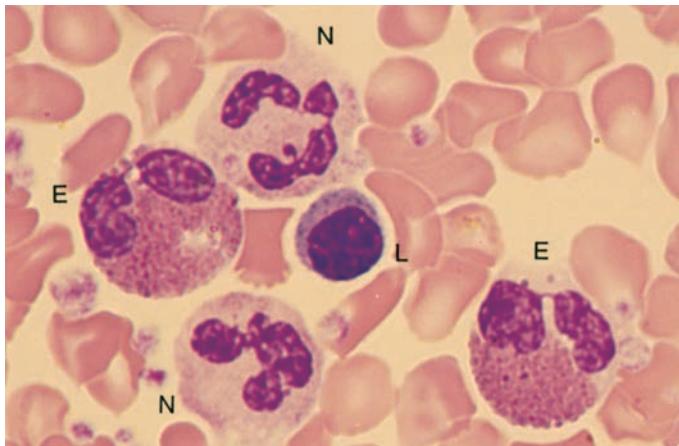


FIGURE 62-26 Normal eosinophils. The film was prepared from the buffy coat of the blood from a normal donor. E, eosinophil; L, lymphocyte; N, neutrophil. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

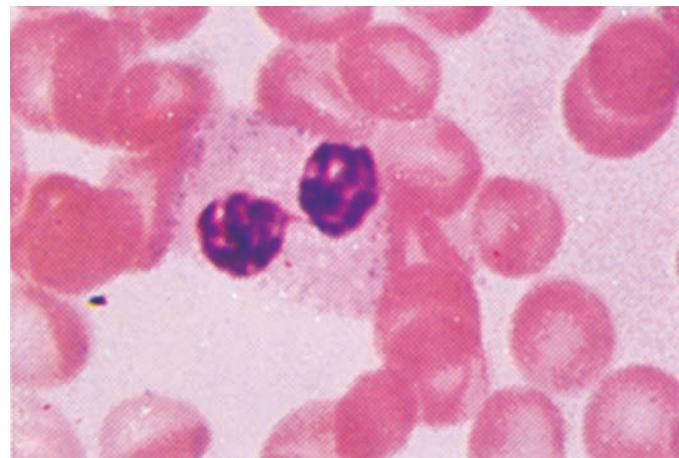


FIGURE 62-28 Pelger-Hüet anomaly. In this benign disorder, the majority of granulocytes are bilobed. The nucleus frequently has a spectacle-like, or "pinçenéz," configuration. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

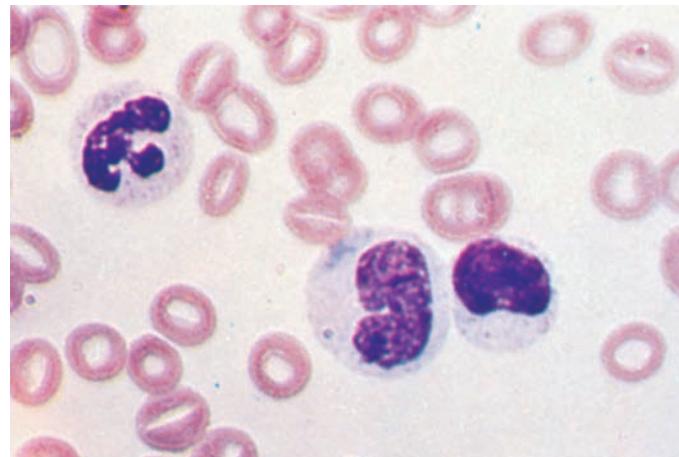


FIGURE 62-29 Döhle body. Neutrophil band with Döhle body. The neutrophil with a sausage-shaped nucleus in the center of the field is a band form. Döhle bodies are discrete, blue-staining nongranular areas found in the periphery of the cytoplasm of the neutrophil in infections and other toxic states. They represent aggregates of rough endoplasmic reticulum. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

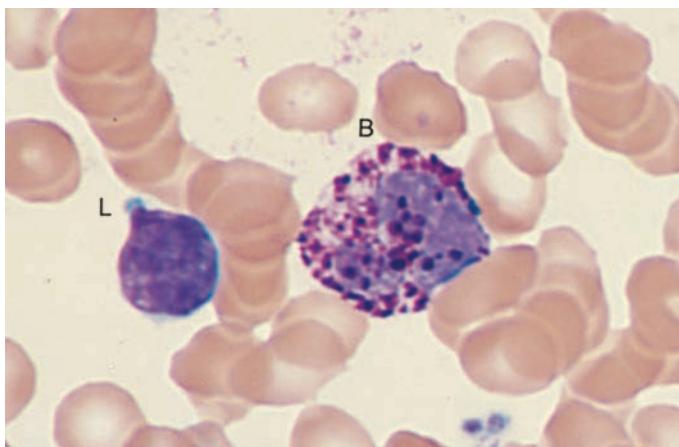


FIGURE 62-27 Normal basophil. The film was prepared from the buffy coat of the blood from a normal donor. B, basophil; L, lymphocyte. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

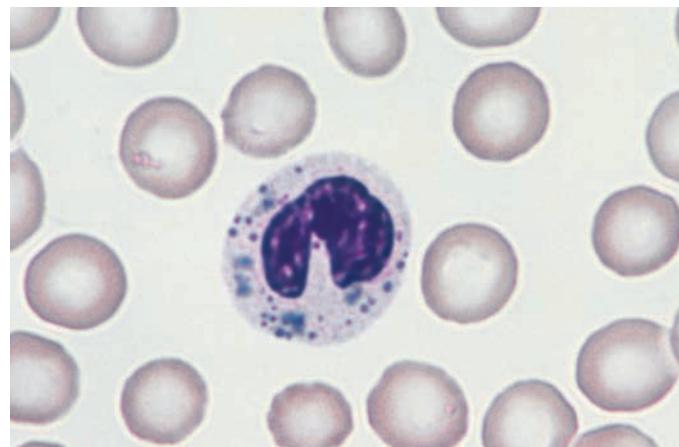


FIGURE 62-30 Chédiak-Higashi disease. Note giant granules in neutrophil. (Source: From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

epithelial malignant cells may be identified. The chances of seeing such abnormal cells are increased by examining blood smears made from buffy coats, the layer of cells that is visible on top of sedimenting red cells when blood is left in the test tube for an hour. Smears made from finger sticks may include rare endothelial cells.

ACKNOWLEDGMENT

Figures in this chapter were borrowed from *Williams Hematology*, 7th edition, M Lichtman et al (eds). New York, McGraw-Hill, 2005; *Hematology in General Practice*, 4th edition, RS Hillman, KA Ault. New York, McGraw-Hill, 2005.

In mammals, O_2 is transported to tissues bound to the hemoglobin contained within circulating red cells. The mature red cell is 8 μm in diameter, anucleate, discoid in shape, and extremely pliable in order to traverse the microcirculation successfully; its membrane integrity is maintained by the intracellular generation of ATP. Normal red cell production results in the daily replacement of 0.8–1% of all circulating red cells in the body, since the average red cell lives 100–120 days. The organ responsible for red cell production is called the *erythron*. The erythron is a dynamic organ made up of a rapidly proliferating pool of marrow erythroid precursor cells and a large mass of mature circulating red blood cells. The size of the red cell mass reflects the balance of red cell production and destruction. The physiologic basis of red cell production and destruction provides an understanding of the mechanisms that can lead to anemia.

The physiologic regulator of red cell production, the glycoprotein hormone EPO, is produced and released by peritubular capillary lining cells within the kidney. These cells are highly specialized epithelial-like cells. A small amount of EPO is produced by hepatocytes. The fundamental stimulus for EPO production is the availability of O_2 for tissue metabolic needs. Key to EPO gene regulation is hypoxia-inducible factor (HIF)-1 α . In the presence of O_2 , HIF-1 α is hydroxylated at a key proline, allowing HIF-1 α to be ubiquitinated and degraded via the proteasome pathway. If O_2 becomes limiting, this critical hydroxylation step does not occur, allowing HIF-1 α to partner with other proteins, translocate to the nucleus, and upregulate the expression of the EPO gene, among others.

Impaired O_2 delivery to the kidney can result from a decreased red cell mass (*anemia*), impaired O_2 loading of the hemoglobin molecule or a high O_2 affinity mutant hemoglobin (*hypoxemia*), or, rarely, impaired blood flow to the kidney (e.g., renal artery stenosis). EPO governs the day-to-day production of red cells, and ambient levels of the hormone can be measured in the plasma by sensitive immunoassays—the normal level being 10–25 U/L. When the hemoglobin concentration falls below 100–120 g/L (10–12 g/dL), plasma EPO levels increase in proportion to the severity of the anemia (Fig. 63-2). In circulation, EPO has a half-clearance time of 6–9 h. EPO acts by binding to specific receptors on the surface of marrow erythroid precursors, inducing them to proliferate and to mature. With EPO stimulation, red cell production can increase four- to fivefold within a 1- to 2-week period, but only in the presence of adequate nutrients, especially iron. The functional capacity of the erythron, therefore, requires normal renal production of EPO, a functioning erythroid marrow, and an adequate supply of substrates for hemoglobin synthesis. A defect in any of these key components can lead to anemia. Generally, anemia is recognized in the laboratory when a patient's hemoglobin level or hematocrit is reduced below an expected value (the normal range).

63

Anemia and Polycythemia

John W. Adamson, Dan L. Longo

HEMATOPOIESIS AND THE PHYSIOLOGIC BASIS OF RED CELL PRODUCTION

Hematopoiesis is the process by which the formed elements of blood are produced. The process is regulated through a series of steps beginning with the hematopoietic stem cell. Stem cells are capable of producing red cells, all classes of granulocytes, monocytes, platelets, and the cells of the immune system. The precise molecular mechanism by which the stem cell becomes committed to a given lineage is not fully defined. However, experiments in mice suggest that erythroid cells come from a common erythroid/megakaryocyte progenitor that does not develop in the absence of expression of the GATA-1 and FOG-1 (friend of GATA-1) transcription factors (Chap. 96). Following lineage commitment, hematopoietic progenitor and precursor cells come increasingly under the regulatory influence of growth factors and hormones. For red cell production, erythropoietin (EPO) is the primary regulatory hormone. EPO is required for the maintenance of committed erythroid progenitor cells that, in the absence of the hormone, undergo programmed cell death (*apoptosis*). The regulated process of red cell production is *erythropoiesis*, and its key elements are illustrated in Fig. 63-1.

In the bone marrow, the first morphologically recognizable erythroid precursor is the pronormoblast. This cell can undergo four to five cell divisions, which result in the production of 16–32 mature red cells. With increased EPO production, or the administration of EPO as a drug, early progenitor cell numbers are amplified and, in turn, give rise to increased numbers of erythrocytes. The regulation of EPO production itself is linked to tissue oxygenation.

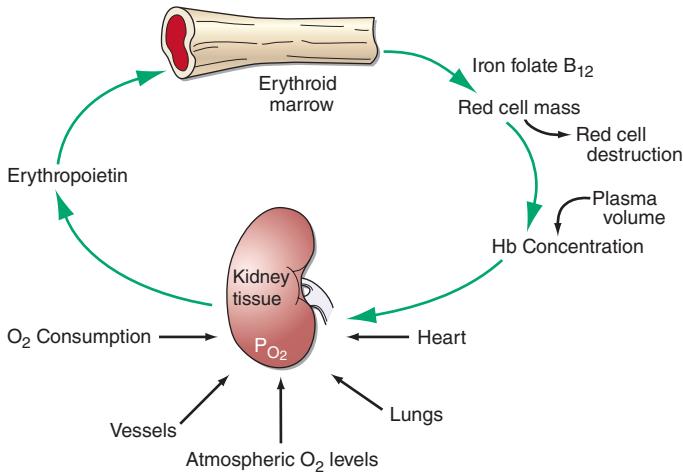


FIGURE 63-1 The physiologic regulation of red cell production by tissue oxygen tension. Hb, hemoglobin.

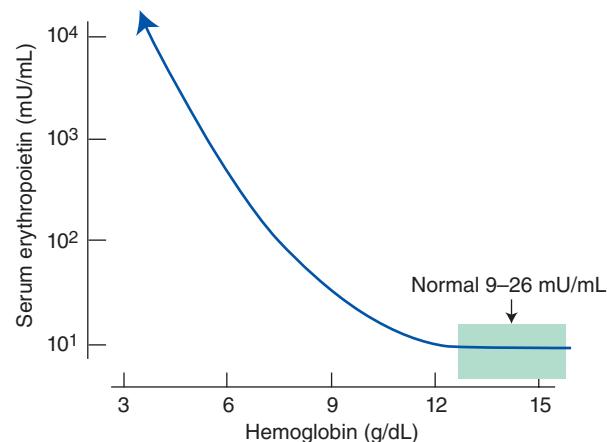


FIGURE 63-2 Erythropoietin (EPO) levels in response to anemia. When the hemoglobin level falls to 120 g/L (12 g/dL), plasma EPO levels increase logarithmically. In the presence of chronic kidney disease or chronic inflammation, EPO levels are typically lower than expected for the degree of anemia. As individuals age, the level of EPO needed to sustain normal hemoglobin levels appears to increase. (Reproduced with permission from RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

The likelihood and severity of anemia are defined based on the deviation of the patient's hemoglobin/hematocrit from values expected for age- and sex-matched normal subjects. The hemoglobin concentration in adults has a Gaussian distribution. The normal range of hemoglobin values for adult males is 13.5–17.5 g/dL (135–175 g/L) and that for adult females is 12–15 g/dL (120–150 g/L). The World Health Organization (WHO) defines anemia as a hemoglobin level <13 g/dL (130 g/L) in men and <12 g/dL (120 g/L) in women. Hematocrit levels are less useful than hemoglobin levels in assessing anemia because they are calculated rather than measured directly. Suspected low hemoglobin or hematocrit values are more easily interpreted if previous values for the same patient are known for comparison.

The critical elements of erythropoiesis—EPO production, iron availability, the proliferative capacity of the bone marrow, and effective maturation of red cell precursors—are used for the initial classification of anemia (see below).

ANEMIA

■ CLINICAL PRESENTATION OF ANEMIA

Signs and Symptoms Anemia is most often recognized by abnormal screening laboratory tests. Patients less commonly present with advanced anemia and its attendant signs and symptoms. Acute anemia is due to blood loss or hemolysis. If blood loss is mild, enhanced O₂ delivery is achieved through changes in the O₂-hemoglobin dissociation curve mediated by a decreased pH or increased CO₂ (*Bohr effect*). With acute blood loss, hypovolemia dominates the clinical picture, and the hematocrit and hemoglobin levels do not reflect the volume of blood lost. Signs of vascular instability appear with acute losses of 10–15% of the total blood volume. In such patients, the issue is not anemia but hypotension and decreased organ perfusion. When >30% of the blood volume is lost suddenly, patients are unable to compensate with the usual mechanisms of vascular contraction and changes in regional blood flow. The patient prefers to remain supine and will show postural hypotension and tachycardia. If the volume of blood lost is >40% (i.e., >2 L in the average-sized adult), signs of hypovolemic shock including confusion, dyspnea, diaphoresis, hypotension, and tachycardia appear (*Chap. 101*). Such patients have significant deficits in vital organ perfusion and require immediate volume replacement.

With acute hemolysis, the signs and symptoms depend on the mechanism that leads to red cell destruction. Intravascular hemolysis with release of free hemoglobin may be associated with acute back pain, free hemoglobin in the plasma and urine, and renal failure. Symptoms associated with more chronic or progressive anemia depend on the age of the patient and the adequacy of blood supply to critical organs. Symptoms associated with moderate anemia include fatigue, loss of stamina, breathlessness, and tachycardia (particularly with physical exertion). However, because of the intrinsic compensatory mechanisms that govern the O₂-hemoglobin dissociation curve, the gradual onset of anemia—particularly in young patients—may not be associated with signs or symptoms until the anemia is severe (hemoglobin <70–80 g/L [7–8 g/dL]). When anemia develops over a period of days or weeks, the total blood volume is normal to slightly increased, and changes in cardiac output and regional blood flow help compensate for the overall loss in O₂-carrying capacity. Changes in the position of the O₂-hemoglobin dissociation curve account for some of the compensatory response to anemia. With chronic anemia, intracellular levels of 2,3-bisphosphoglycerate rise, shifting the dissociation curve to the right and facilitating O₂ unloading. This compensatory mechanism can only maintain normal tissue O₂ delivery in the face of a 20–30 g/L (2–3 g/dL) deficit in hemoglobin concentration. Finally, further protection of O₂ delivery to vital organs is achieved by the shunting of blood away from organs that are relatively rich in blood supply, particularly the kidney, gut, and skin.

Certain disorders are commonly associated with anemia. Chronic inflammatory states (e.g., infection, rheumatoid arthritis, cancer) are associated with mild to moderate anemia, whereas lymphoproliferative disorders, such as chronic lymphocytic leukemia and certain other B-cell neoplasms, may be associated with autoimmune hemolysis.

APPROACH TO THE PATIENT

Anemia

The evaluation of the patient with anemia requires a careful history and physical examination. Nutritional history related to drugs or alcohol intake and family history of anemia should always be assessed. Certain geographic backgrounds and ethnic origins are associated with an increased likelihood of an inherited disorder of the hemoglobin molecule or intermediary metabolism. Glucose-6-phosphate dehydrogenase (G6PD) deficiency and certain hemoglobinopathies are seen more commonly in those of Middle Eastern or African origin, including blacks who have a high frequency of G6PD deficiency. Other information that may be useful includes exposure to certain toxic agents or drugs and symptoms related to other disorders commonly associated with anemia. These include symptoms and signs such as bleeding, fatigue, malaise, fever, weight loss, night sweats, and other systemic symptoms. Clues to the mechanisms of anemia may be provided on physical examination by findings of infection, blood in the stool, lymphadenopathy, splenomegaly, or petechiae. Splenomegaly and lymphadenopathy suggest an underlying lymphoproliferative disease, whereas petechiae suggest platelet dysfunction. Past laboratory measurements are helpful to determine a time of onset.

In the anemic patient, physical examination may demonstrate a forceful heartbeat, strong peripheral pulses, and a systolic “flow” murmur. The skin and mucous membranes may be pale if the hemoglobin is <8–10 g/dL (80–100 g/L). This part of the physical examination should focus on areas where vessels are close to the surface such as the mucous membranes, nail beds, and palmar creases. If the palmar creases are lighter in color than the surrounding skin when the hand is hyperextended, the hemoglobin level is usually <8 g/dL (80 g/L).

LABORATORY EVALUATION

Table 63-1 lists the tests used in the initial workup of anemia. A routine complete blood count (CBC) is required as part of the evaluation and includes the hemoglobin, hematocrit, and red cell indices: the mean cell volume (MCV) in femtoliters, mean cell hemoglobin (MCH) in picograms per cell, and mean concentration of hemoglobin per volume of red cells (MCHC) in grams per liter (non-SI: grams per deciliter). The MCH is the least useful of the indices; it tends to track with the MCV. The red cell indices are calculated as shown in **Table 63-2**, and the normal variations in the hemoglobin and hematocrit with age are shown in **Table 63-3**. A number of physiologic factors affect the CBC, including age, sex, pregnancy, smoking, and altitude. High-normal hemoglobin values may be seen in men and women who live at altitude or smoke heavily. Hemoglobin elevations due to smoking reflect normal compensation due to the displacement of O₂ by CO in hemoglobin binding. Other important information is provided by the reticulocyte count and measurements of iron supply including *serum iron*, *total iron-binding capacity* (TIBC; an indirect measure of serum transferrin), and *serum ferritin*. Marked alterations in the red cell indices usually reflect disorders of maturation or iron deficiency. A careful evaluation of the peripheral blood smear is important, and clinical laboratories often provide a description of both the red and white cells, a white cell differential count, and the platelet count. In patients with severe anemia and abnormalities in red blood cell morphology and/or low reticulocyte counts, a bone marrow aspirate or biopsy can assist in the diagnosis. Other tests of value in the diagnosis of specific anemias are discussed in chapters on specific disease states.

The components of the CBC also help in the classification of anemia. *Microcytosis* is reflected by a lower than normal MCV (<80), whereas high values (>100) reflect *macrocytosis*. The MCHC reflects defects in hemoglobin synthesis (*hypochromia*). Automated cell counters describe the red cell volume distribution width (RDW). The MCV (representing the peak of the distribution curve)

TABLE 63-1 Laboratory Tests in Anemia Diagnosis

- I. Complete blood count (CBC)
 - A. Red blood cell count
 - 1. Hemoglobin
 - 2. Hematocrit
 - 3. Reticulocyte count
 - B. Red blood cell indices
 - 1. Mean cell volume (MCV)
 - 2. Mean cell hemoglobin (MCH)
 - 3. Mean cell hemoglobin concentration (MCHC)
 - 4. Red cell distribution width (RDW)
 - C. White blood cell count
 - 1. Cell differential
 - 2. Nuclear segmentation of neutrophils
 - D. Platelet count
 - E. Cell morphology
 - 1. Cell size
 - 2. Hemoglobin content
 - 3. Anisocytosis
 - 4. Poikilocytosis
 - 5. Polychromasia
- II. Iron supply studies
 - A. Serum iron
 - B. Total iron-binding capacity
 - C. Serum ferritin
- III. Marrow examination
 - A. Aspirate
 - 1. M/E ratio^a
 - 2. Cell morphology
 - 3. Iron stain
 - B. Biopsy
 - 1. Cellularity
 - 2. Morphology

^aM/E ratio, ratio of myeloid to erythroid precursors.

TABLE 63-2 Red Blood Cell Indices

INDEX	NORMAL VALUE
Mean cell volume (MCV) = (hematocrit × 10)/ (red cell count × 10 ⁶)	90 ± 8 fL
Mean cell hemoglobin (MCH) = (hemoglobin × 10)/ (red cell count × 10 ⁶)	30 ± 3 pg
Mean cell hemoglobin concentration = (hemoglobin × 10)/hematocrit, or MCH/MCV	33 ± 2%

TABLE 63-3 Changes in Normal Hemoglobin/Hematocrit Values with Age, Sex, and Pregnancy

AGE/SEX	HEMOGLOBIN, g/dL	HEMATOCRIT, %
At birth	17	52
Childhood	12	36
Adolescence	13	40
Adult man	16 (±2)	47 (±6)
Adult woman (menstruating)	13 (±2)	40 (±6)
Adult woman (postmenopausal)	14 (±2)	42 (±6)
During pregnancy	12 (±2)	37 (±6)

Source: From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.

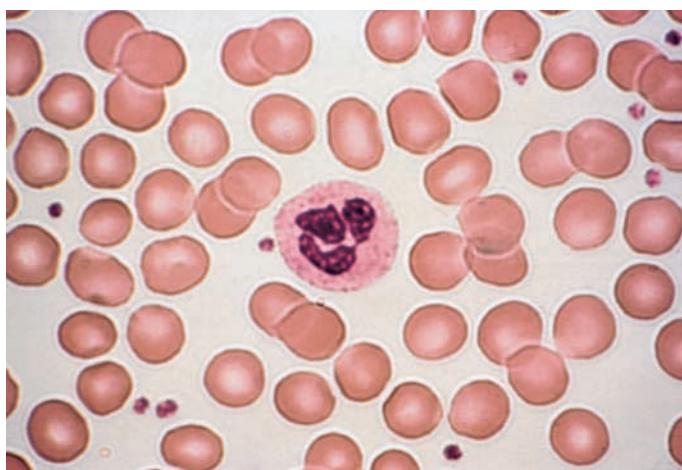


FIGURE 63-3 Normal blood smear (Wright stain). High-power field showing normal red cells, a neutrophil, and a few platelets. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

is insensitive to the appearance of small populations of macrocytes or microcytes. An experienced laboratory technician will be able to identify minor populations of large or small cells or hypochromic cells on the peripheral blood film before the red cell indices change.

Peripheral Blood Smear The peripheral blood smear provides important information about defects in red cell production (Chap. 62). As a complement to the red cell indices, the blood smear also reveals variations in cell size (*anisocytosis*) and shape (*poikilocytosis*). The degree of anisocytosis usually correlates with increases in the RDW or the range of cell sizes. Poikilocytosis suggests a defect in the maturation of red cell precursors in the bone marrow or fragmentation of circulating red cells. The blood smear may also reveal *polychromasia*—red cells that are slightly larger than normal and grayish blue in color on the Wright-Giemsa stain. These cells are reticulocytes that have been released prematurely from the bone marrow and their color represents residual amounts of ribosomal RNA. These cells appear in circulation in response to EPO stimulation or to architectural damage of the bone marrow (fibrosis, infiltration of the marrow by malignant cells, etc.) that results in their disordered release from the marrow. The appearance of nucleated red cells, Howell-Jolly bodies, target cells, sickle cells, and other changes may provide clues to specific disorders (Figs. 63-3 to 63-11).

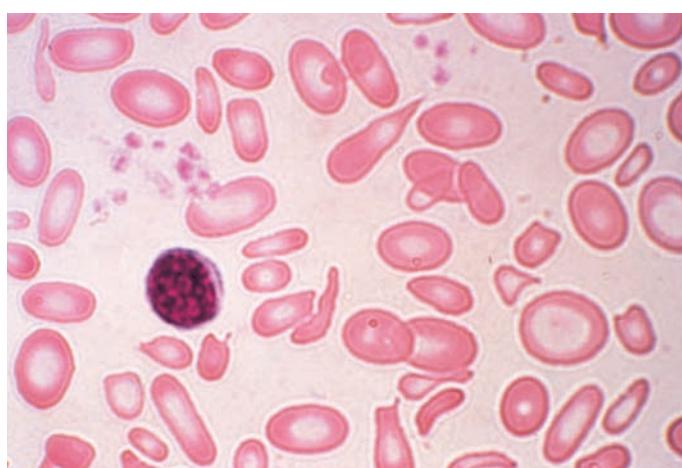


FIGURE 63-4 Severe iron-deficiency anemia. Microcytic and hypochromic red cells smaller than the nucleus of a lymphocyte associated with marked variation in size (anisocytosis) and shape (poikilocytosis). (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

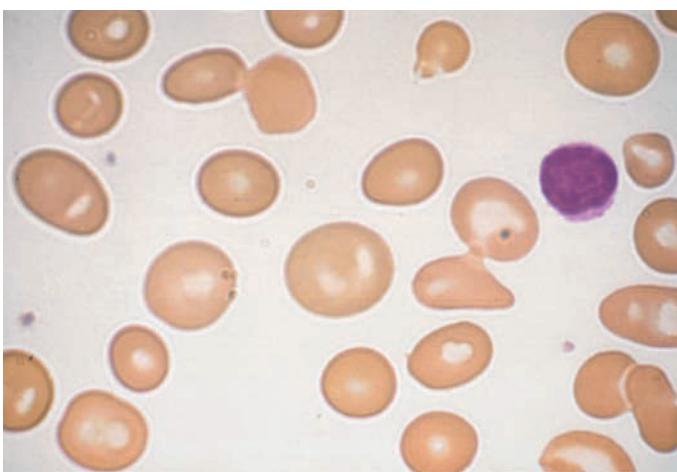


FIGURE 63-5 Macrocytosis. Red cells are larger than a small lymphocyte and well hemoglobinized. Often macrocytes are oval shaped (macro-ovalocytes). (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

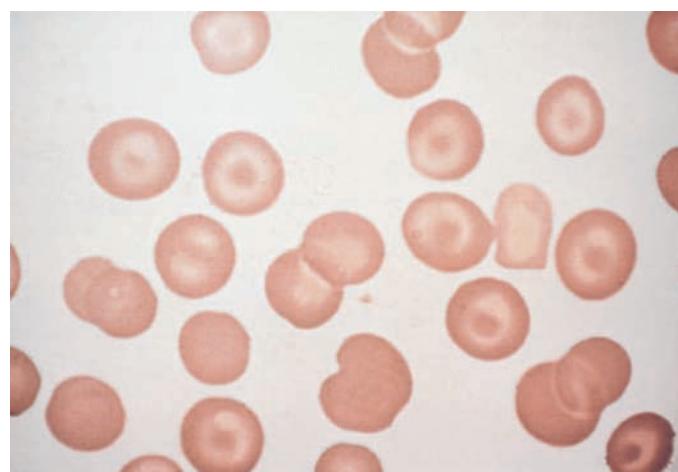


FIGURE 63-8 Target cells. Target cells have a bull's-eye appearance and are seen in thalassemia and in liver disease. (From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

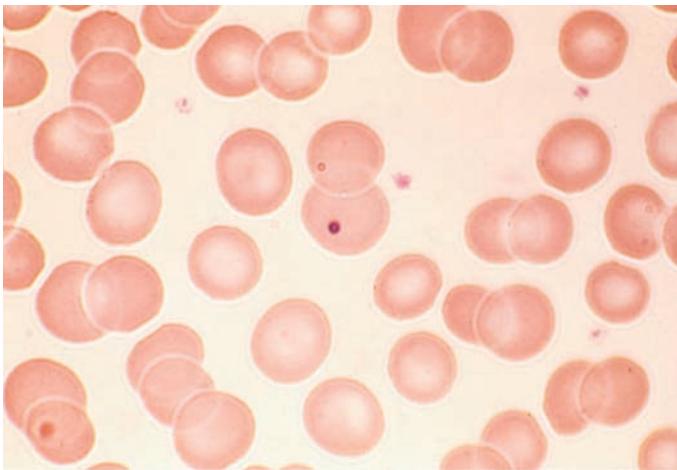


FIGURE 63-6 Howell-Jolly bodies. In the absence of a functional spleen, nuclear remnants are not culled from the red cells and remain as small homogeneously staining blue inclusions on Wright stain. (From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw-Hill, 2005.)

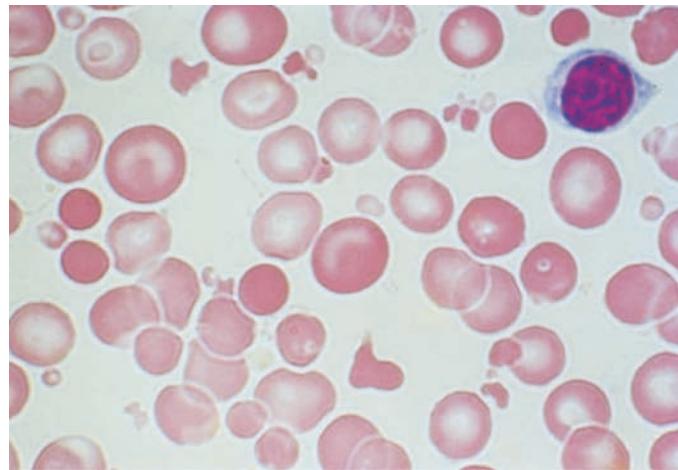


FIGURE 63-9 Red cell fragmentation. Red cells may become fragmented in the presence of foreign bodies in the circulation, such as mechanical heart valves, or in the setting of thermal injury. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

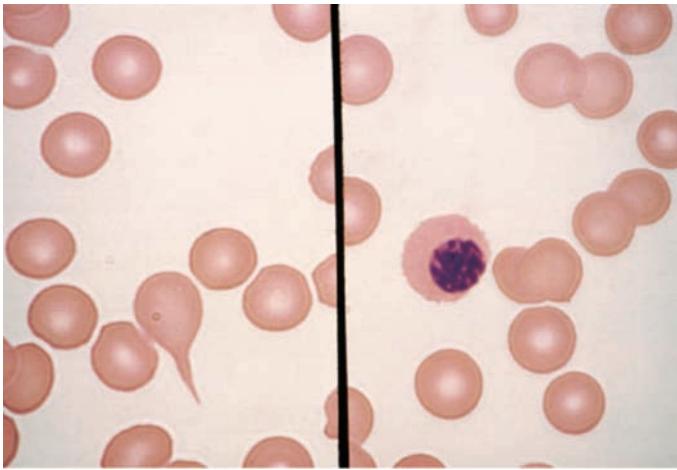


FIGURE 63-7 Red cell changes in myelofibrosis. The left panel shows a teardrop-shaped cell. The right panel shows a nucleated red cell. These forms can be seen in myelofibrosis. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

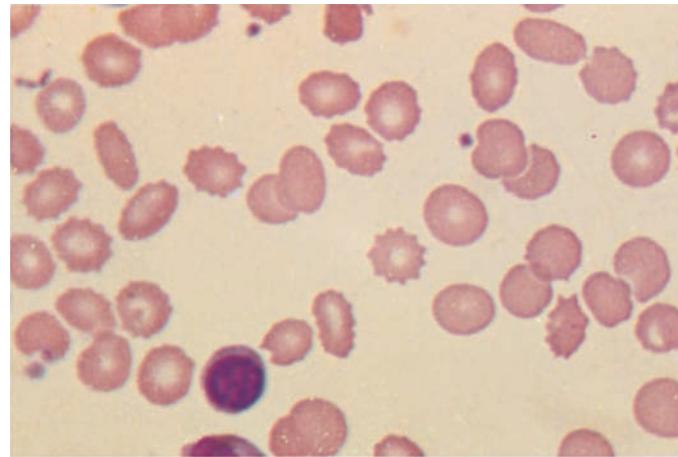


FIGURE 63-10 Uremia. The red cells in uremia may acquire numerous regularly spaced, small, spiny projections. Such cells, called burr cells or echinocytes, are readily distinguishable from irregularly spiculated acanthocytes shown in Fig. 63-11. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

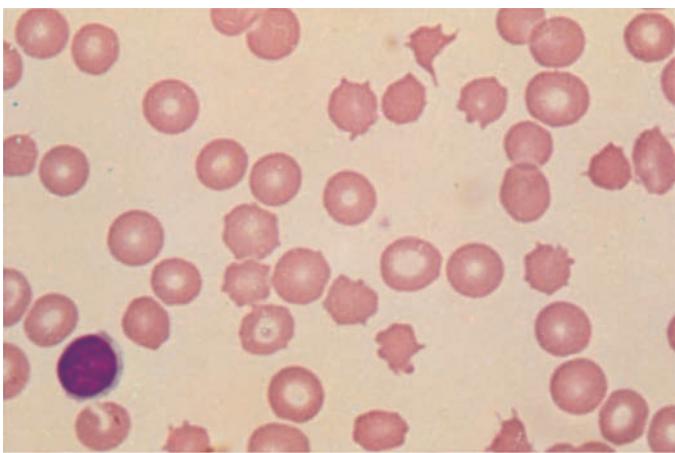


FIGURE 63-11 Spur cells. Spur cells are recognized as distorted red cells containing several irregularly distributed thorn-like projections. Cells with this morphologic abnormality are also called acanthocytes. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

Reticulocyte Count An accurate reticulocyte count is key to the initial classification of anemia. Reticulocytes are red cells that have been recently released from the bone marrow. They are identified by staining with a supravital dye that precipitates the ribosomal RNA (Fig. 63-12). These precipitates appear as blue or black punctate spots and can be counted manually or, currently, by fluorescent emission of dyes that bind to RNA. This residual RNA is metabolized over the first 24–36 h of the reticulocyte's life span in circulation. Normally, the reticulocyte count ranges from 1% to 2% and reflects the daily replacement of 0.8–1.0% of the circulating red cell population. A corrected reticulocyte percentage or the absolute number of reticulocytes provides a reliable measure of effective red cell production.

In the initial classification of anemia, the patient's reticulocyte count is compared with the expected reticulocyte response. In general, if the EPO and erythroid marrow responses to moderate anemia [hemoglobin <100 g/L (10 g/dL)] are intact, the red cell production rate increases to two to three times normal within 10 days following the onset of anemia. In the face of established anemia, a reticulocyte response less than two to three times normal indicates an inadequate marrow response.

To use the reticulocyte count to estimate marrow response, two corrections are necessary. The first correction adjusts the reticulocyte count based on the reduced number of circulating red cells. With anemia, the percentage of reticulocytes may be increased

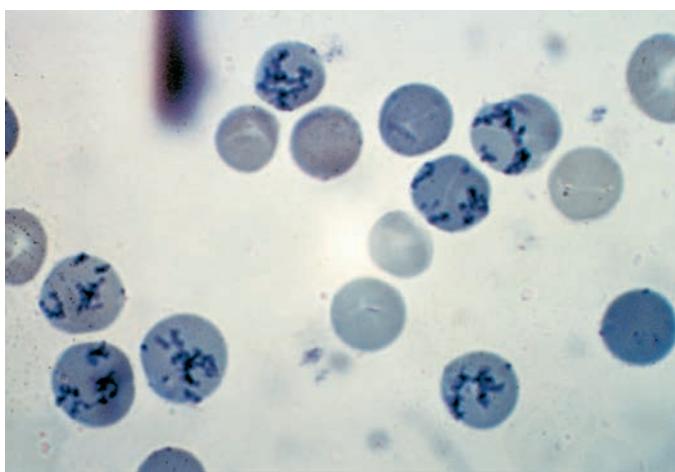


FIGURE 63-12 Reticulocytes. Methylene blue stain demonstrates residual RNA in newly made red cells. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

TABLE 63-4 Calculation of Reticulocyte Production Index

Correction #1 for Anemia:

This correction produces the corrected reticulocyte count.

In a person whose reticulocyte count is 9%, hemoglobin 7.5 g/dL, and hematocrit 23%, the absolute reticulocyte count = $9 \times (7.5/15)$ [or $\times (23/45)$] = 4.5%

Note. This correction is not done if the reticulocyte count is reported in absolute numbers (e.g., 50,000/ μ L of blood)

Correction #2 for Longer Life of Prematurely Released Reticulocytes in the Blood:

This correction produces the reticulocyte production index.

In a person whose reticulocyte count is 9%, hemoglobin 7.5 g/dL, and hematocrit 23%, the reticulocyte production index

$$= 9 \times \frac{(7.5/15)(\text{hemoglobin correction})}{2(\text{maturation time correction})} = 2.25$$

while the absolute number is unchanged. To correct for this effect, the reticulocyte percentage is multiplied by the ratio of the patient's hemoglobin or hematocrit to the expected hemoglobin/hematocrit for the age and sex of the patient (Table 63-4). This provides an estimate of the reticulocyte count corrected for anemia. To convert the corrected reticulocyte count to an index of marrow production, a further correction is required, depending on whether some of the reticulocytes in circulation have been released from the marrow prematurely. For this second correction, the peripheral blood smear is examined to see if there are polychromatophilic macrocytes present.

These cells, representing prematurely released reticulocytes, are referred to as "shift" cells, and the relationship between the degree of shift and the necessary shift correction factor is shown in Fig. 63-13. The correction is necessary because these prematurely released cells survive as reticulocytes in circulation for >1 day, thereby providing a falsely high estimate of daily red cell production. If polychromasia is increased, the reticulocyte count, already corrected for anemia, should be corrected again by 2 to account for the prolonged reticulocyte maturation time. The second correction factor varies from 1 to 3 depending on the severity of anemia. To simplify things, a correction of 2 is used. An appropriate correction is shown in Table 63-4. If polychromatophilic cells are not seen on

Hematocrit (%)	Marrow normoblasts and reticulocytes (days)	Peripheral blood reticulocytes (days)
45	3.5	1.0
35	3.0	1.5
25	2.5	2.0
15	1.5	2.5

↓

"SHIFT" correction factor

FIGURE 63-13 Correction of the reticulocyte count. To use the reticulocyte count as an indicator of effective red cell production, the reticulocyte number must be corrected based on the level of anemia and the circulating life span of the reticulocytes. Erythroid cells take ~4.5 days to mature. At a normal hemoglobin, reticulocytes are released to the circulation with ~1 day left as reticulocytes. However, with different levels of anemia, reticulocytes (and even earlier erythroid cells) may be released from the marrow prematurely. Most patients come to clinical attention with hematocrits in the mid-20s, and thus a correction factor of 2 is commonly used because the observed reticulocytes will live for 2 days in the circulation before losing their RNA.

TABLE 63-5 Normal Marrow Response to Anemia

HEMOGLOBIN	PRODUCTION INDEX	RETICULOCYTE COUNT
15 g/dL	1	50,000/ μ L
11 g/dL	2.0–2.5	100–150,000/ μ L
8 g/dL	3.0–4.0	300–400,000/ μ L

the blood smear, the second correction is not indicated. The now doubly corrected reticulocyte count is the *reticulocyte production index*, and it provides an estimate of marrow production relative to normal. In many hospital laboratories, the reticulocyte count is reported not only as a percentage but also in absolute numbers. If so, no correction for dilution is required. A summary of the appropriate marrow response to varying degrees of anemia is shown in **Table 63-5**.

Premature release of reticulocytes is normally due to increased EPO stimulation. However, if the integrity of the bone marrow release process is lost through tumor infiltration, fibrosis, or other disorders, the appearance of nucleated red cells or polychromatophilic macrocytes should still invoke the second reticulocyte correction. The shift correction should always be applied to a patient with anemia and a very high reticulocyte count to provide a true index of effective red cell production. Patients with severe chronic hemolytic anemia may increase red cell production as much as six- to sevenfold. This measure alone confirms the fact that the patient has an appropriate EPO response, a normally functioning bone marrow, and sufficient iron available to meet the demands for new red cell formation. If the reticulocyte production index is <2 in the face of established anemia, a defect in erythroid marrow proliferation or maturation must be present.

Tests of Iron Supply and Storage The laboratory measurements that reflect the availability of iron for hemoglobin synthesis include the serum iron, the TIBC, and the percent transferrin saturation. The percent transferrin saturation is derived by dividing the serum iron level ($\times 100$) by the TIBC. The normal serum iron ranges from 9 to 27 μ mol/L (50–150 μ g/dL), whereas the normal TIBC is 54–64 μ mol/L (300–360 μ g/dL); the normal transferrin saturation ranges from 25 to 50%. A diurnal variation in the serum iron leads to a variation in the percent transferrin saturation. The serum ferritin is used to evaluate total body iron stores. Adult males have serum ferritin levels that average $\sim 100 \mu\text{g/L}$, corresponding to iron stores of $\sim 1 \text{ g}$. Adult premenopausal females have lower serum ferritin levels averaging 30 $\mu\text{g/L}$, reflecting lower iron stores ($\sim 300 \text{ mg}$). A serum ferritin level of 10–15 $\mu\text{g/L}$ indicates depletion of body iron stores. However, ferritin is also an acute-phase reactant and, in the presence of acute or chronic inflammation, may rise several-fold above baseline levels. As a rule, a serum ferritin $>200 \mu\text{g/L}$ means there is at least some iron in tissue stores.

Bone Marrow Examination A bone marrow aspirate and smear or a needle biopsy can be useful in the evaluation of some patients with anemia. In patients with hypoproliferative anemia, normal renal function, and normal iron status, a bone marrow is indicated. Marrow examination can diagnose primary marrow disorders such as myelofibrosis, a red cell maturation defect, or an infiltrative disease (**Figs. 63-14 to 63-16**). The increase or decrease of one cell lineage (myeloid vs erythroid) compared to another is obtained by a differential count of nucleated cells in a bone marrow smear (the myeloid/erythroid [M/E] ratio). A patient with a hypoproliferative anemia (see below) and a reticulocyte production index <2 will demonstrate an M/E ratio of 2 or 3:1. In contrast, patients with hemolytic disease and a production index >3 will have an M/E ratio of at least 1:1. Maturation disorders are identified from the discrepancy between the M/E ratio and the reticulocyte production index (see below). Either the marrow smear or biopsy can be stained for the presence of iron stores or iron in developing red cells. The

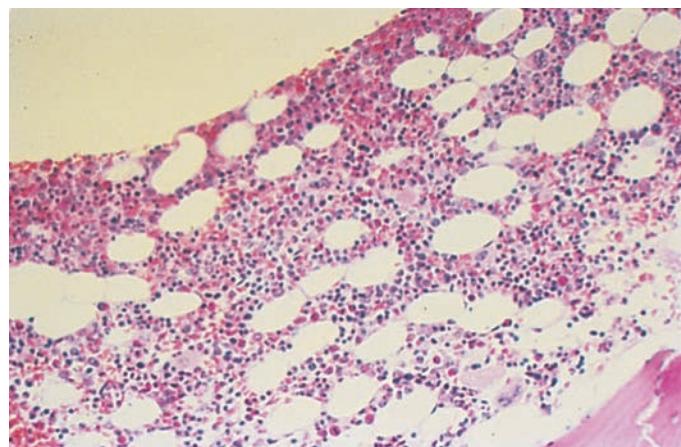


FIGURE 63-14 Normal bone marrow. This is a low-power view of a section of a normal bone marrow biopsy stained with hematoxylin and eosin (H&E). Note that the nucleated cellular elements account for $\sim 40\text{--}50\%$ and the fat (clear areas) accounts for $\sim 50\text{--}60\%$ of the area. (From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)

storage iron is in the form of ferritin or *hemosiderin*. On carefully prepared bone marrow smears, small ferritin granules can normally be seen under oil immersion in 20–40% of developing erythroblasts. Such cells are called *sideroblasts*.

OTHER LABORATORY MEASUREMENTS

Additional laboratory tests may be of value in confirming specific diagnoses. **For details of these tests and how they are applied in individual disorders, see Chaps. 97 to 101.**

DEFINITION AND CLASSIFICATION OF ANEMIA

Initial Classification of Anemia The functional classification of anemia has three major categories. These are (1) marrow production defects (*hypoproliferation*), (2) red cell maturation defects (*ineffective erythropoiesis*), and (3) decreased red cell survival (*blood loss/hemolysis*). The classification is shown in **Fig. 63-17**. A hypoproliferative anemia is typically seen with a low reticulocyte production index together with little or no change in red cell morphology (a normocytic, normochromic anemia) (**Chap. 97**). Maturation disorders typically have a slight to moderately elevated reticulocyte production index that is accompanied by either macrocytic (**Chap. 99**) or microcytic (**Chaps. 97, 98**) red cell

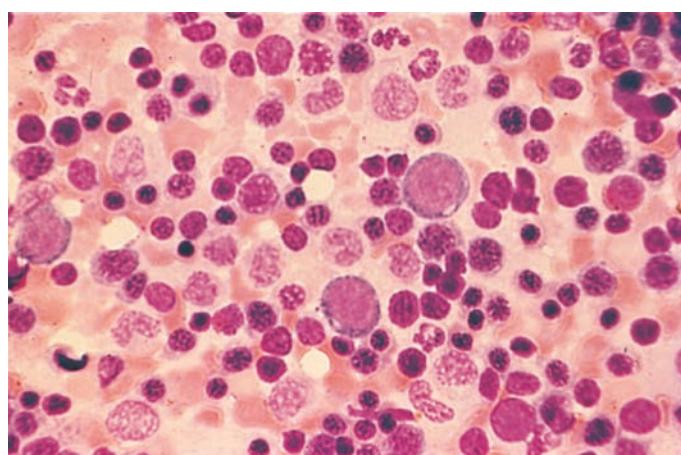


FIGURE 63-15 Erythroid hyperplasia. This marrow shows an increase in the fraction of cells in the erythroid lineage as might be seen when a normal marrow compensates for acute blood loss or hemolysis. The myeloid/erythroid (M/E) ratio is about 1:1. (From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)

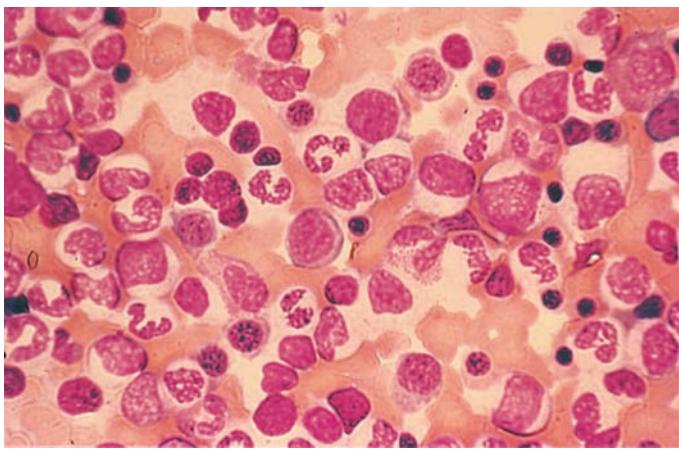


FIGURE 63-16 Myeloid hyperplasia. This marrow shows an increase in the fraction of cells in the myeloid or granulocytic lineage as might be seen in a normal marrow responding to infection. The myeloid/erythroid (M/E) ratio is $>3:1$. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

indices. Increased red blood cell destruction secondary to hemolysis results in an increase in the reticulocyte production index to at least three times normal (Chap. 100), provided sufficient iron is available. Hemorrhagic anemia does not typically result in production indices of more than 2.0–2.5 times normal because of the limitations placed on expansion of the erythroid marrow by iron availability (Chap. 101).

In the first branch point of the classification of anemia, a reticulocyte production index >2.5 indicates that hemolysis is most likely. A reticulocyte production index <2 indicates either a hypoproliferative anemia or maturation disorder. The latter two possibilities can often be distinguished by the red cell indices, by examination of the peripheral blood smear, or by a marrow examination. If the red cell indices are normal, the anemia is almost certainly hypoproliferative

in nature. Maturation disorders are characterized by ineffective red cell production and a low reticulocyte production index. Bizarre red cell shapes—macrocytes or hypochromic microcytes—are seen on the peripheral blood smear. With a hypoproliferative anemia, no erythroid hyperplasia is noted in the marrow, whereas patients with ineffective red cell production have erythroid hyperplasia and an M/E ratio $<1:1$.

Hypoproliferative Anemias At least 75% of all cases of anemia are hypoproliferative in nature. A hypoproliferative anemia reflects absolute or relative marrow failure in which the erythroid marrow has not proliferated appropriately for the degree of anemia. The majority of hypoproliferative anemias are due to mild to moderate iron deficiency or inflammation. A hypoproliferative anemia can result from marrow damage, iron deficiency, or inadequate EPO stimulation. The last may reflect impaired renal function, suppression of EPO production by inflammatory cytokines such as interleukin 1, or reduced tissue needs for O_2 from metabolic disease such as hypothyroidism. Only occasionally is the marrow unable to produce red cells at a normal rate, and this is most prevalent in patients with renal failure. With diabetes mellitus or myeloma, the EPO deficiency may be more marked than would be predicted by the degree of renal insufficiency. In general, hypoproliferative anemias are characterized by normocytic, normochromic red cells, although microcytic, hypochromic cells may be observed with mild iron deficiency or long-standing chronic inflammatory disease. The key laboratory tests in distinguishing between the various forms of hypoproliferative anemia include the serum iron and iron-binding capacity, evaluation of renal and thyroid function, a marrow biopsy or aspirate to detect marrow damage or infiltrative disease, and serum ferritin to assess iron stores. An iron stain of the marrow will determine the pattern of iron distribution. Patients with the anemia of acute or chronic inflammation show a distinctive pattern of serum iron (low), TIBC (normal or low), percent transferrin saturation (low), and serum ferritin (normal or high). These changes in iron values are brought about by hepcidin, the iron regulatory hormone that is produced by the liver and is increased in inflammation (Chap. 97). A distinct pattern of results is noted in mild to moderate iron deficiency (low serum iron, high TIBC, low percent transferrin saturation, low serum ferritin) (Chap. 97). Marrow damage by drugs, infiltrative disease such as leukemia or lymphoma, or marrow aplasia is diagnosed from the peripheral blood and bone marrow morphology. With infiltrative disease or fibrosis, a marrow biopsy is required.

Maturation Disorders The presence of anemia with an inappropriately low reticulocyte production index, macro- or microcytosis on smear, and abnormal red cell indices suggests a maturation disorder. Maturation disorders are divided into two categories: nuclear maturation defects, associated with macrocytosis, and cytoplasmic maturation defects, associated with microcytosis and hypochromia usually from defects in hemoglobin synthesis. The inappropriately low reticulocyte production index is a reflection of the ineffective erythropoiesis that results from the destruction within the marrow of developing erythroblasts. Bone marrow examination shows erythroid hyperplasia.

Nuclear maturation defects result from vitamin B₁₂ or folic acid deficiency, drug damage, or myelodysplasia. Drugs that interfere with cellular DNA synthesis, such as methotrexate or alkylating agents, can produce a nuclear maturation defect. Alcohol, alone, is also capable of producing macrocytosis and a variable degree of anemia, but this is usually associated with folic acid deficiency. Measurements of folic acid and vitamin B₁₂ are critical not only in identifying the specific vitamin deficiency but also because they reflect different pathogenetic mechanisms (Chap. 99).

Cytoplasmic maturation defects result from severe iron deficiency or abnormalities in globin or heme synthesis. Iron deficiency occupies an unusual position in the classification of anemia. If the iron-deficiency anemia is mild to moderate, erythroid marrow proliferation is blunted and the anemia is classified as hypoproliferative. However, if the anemia is severe and prolonged, the erythroid marrow will become hyperplastic despite the inadequate iron supply, and the anemia will be classified as ineffective erythropoiesis with a cytoplasmic maturation defect. In either case, an inappropriately low reticulocyte

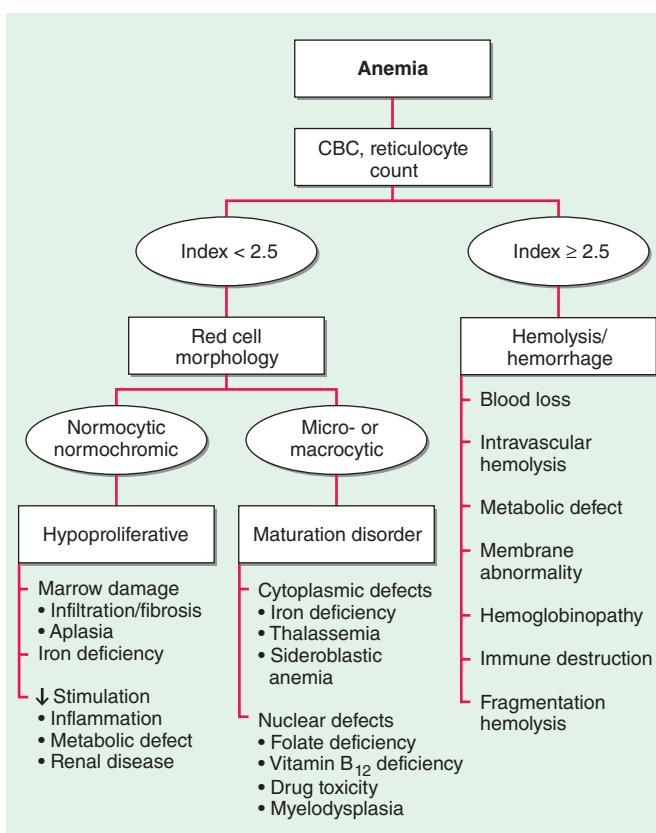


FIGURE 63-17 The physiologic classification of anemia. CBC, complete blood count.

production index, microcytosis, and a classic pattern of iron values make the diagnosis clear and easily distinguish iron deficiency from other cytoplasmic maturation defects such as the thalassemias. Defects in heme synthesis, in contrast to globin synthesis, are less common and may be acquired or inherited (Chap. 416). Acquired abnormalities are usually associated with myelodysplasia, may lead to either a macro- or microcytic anemia, and are frequently associated with mitochondrial iron loading. In these cases, iron is taken up by the mitochondria of the developing erythroid cell but not incorporated into heme. The iron-encrusted mitochondria surround the nucleus of the erythroid cell, forming a ring. Based on the distinctive finding of so-called ringed sideroblasts on the marrow iron stain, patients are diagnosed as having a sideroblastic anemia—almost always reflecting myelodysplasia. Again, studies of iron parameters are helpful in the differential diagnosis of these patients.

Blood Loss/Hemolytic Anemia In contrast to anemias associated with an inappropriately low reticulocyte production index, hemolysis is associated with red cell production indices ≥ 2.5 times normal. The stimulated erythropoiesis is reflected in the blood smear by the appearance of increased numbers of polychromatophilic macrocytes. A marrow examination is rarely indicated if the reticulocyte production index is increased appropriately. The red cell indices are typically normocytic or slightly macrocytic, reflecting the increased number of reticulocytes. Acute blood loss is not associated with an increased reticulocyte production index because of the time required to increase EPO production and, subsequently, marrow proliferation (Chap. 101). Subacute blood loss may be associated with modest reticulocytosis. Anemia from chronic blood loss presents more often as iron deficiency than with the picture of increased red cell production.

The evaluation of blood loss anemia is usually not difficult. Most problems arise when a patient presents with an increased red cell production index from an episode of acute blood loss that went unrecognized. The cause of the anemia and increased red cell production may not be obvious. The confirmation of a recovering state may require observations over a period of 2–3 weeks, during which the hemoglobin concentration will rise and the reticulocyte production index fall (Chap. 101).

Hemolytic disease, while dramatic, is among the least common forms of anemia. The ability to sustain a high reticulocyte production index reflects the ability of the erythroid marrow to compensate for hemolysis and, in the case of extravascular hemolysis, the efficient recycling of iron from the destroyed red cells to support red cell production. With intravascular hemolysis, such as paroxysmal nocturnal hemoglobinuria, the loss of iron may limit the marrow response. The level of response depends on the severity of the anemia and the nature of the underlying disease process.

Hemoglobinopathies, such as sickle cell disease and the thalassemias, present a mixed picture. The reticulocyte index may be high but is inappropriately low for the degree of marrow erythroid hyperplasia (Chap. 98).

Hemolytic anemias present in different ways. Some appear suddenly as an acute, self-limited episode of intravascular or extravascular hemolysis, a presentation pattern often seen in patients with autoimmune hemolysis or with inherited defects of the Embden-Meyerhof pathway or the glutathione reductase pathway. Patients with inherited disorders of the hemoglobin molecule or red cell membrane generally have a lifelong clinical history typical of the disease process. Those with chronic hemolytic disease, such as hereditary spherocytosis, may actually present not with anemia but with a complication stemming from the prolonged increase in red cell destruction such as symptomatic bilirubin gallstones or splenomegaly. Patients with chronic hemolysis are also susceptible to aplastic crises if an infectious process interrupts red cell production.

The differential diagnosis of an acute or chronic hemolytic event requires the careful integration of family history, the pattern of clinical presentation, and—whether the disease is congenital or acquired—careful examination of the peripheral blood smear. Precise diagnosis may require more specialized laboratory tests, such as hemoglobin

electrophoresis or a screen for red cell enzymes. Acquired defects in red cell survival are often immunologically mediated and require a direct or indirect antiglobulin test or a cold agglutinin titer to detect the presence of hemolytic antibodies or complement-mediated red cell destruction (Chap. 100).

TREATMENT

Anemia

An overriding principle is to initiate treatment of mild to moderate anemia only when a specific diagnosis is made. Rarely, in the acute setting, anemia may be so severe that red cell transfusions are required before a specific diagnosis is available. Whether the anemia is of acute or gradual onset, the selection of the appropriate treatment is determined by the documented cause(s) of the anemia. Often, the cause of the anemia is multifactorial. For example, a patient with severe rheumatoid arthritis who has been taking anti-inflammatory drugs may have a hypoproliferative anemia associated with chronic inflammation as well as chronic blood loss associated with intermittent gastrointestinal bleeding. In every circumstance, it is important to evaluate the patient's iron status fully before and during the treatment of any anemia. Transfusion is discussed in Chap. 113; iron therapy is discussed in Chap. 97; treatment of megaloblastic anemia is discussed in Chap. 99; treatment of other entities is discussed in their respective chapters (sickle cell anemia, Chap. 98; megaloblastic anemia, Chap. 99; hemolytic anemias, Chap. 100; aplastic anemia and myelodysplasia, Chap. 102).

Therapeutic options for the treatment of anemias have expanded dramatically during the past 30 years. Blood component therapy is available and safe. Recombinant EPO as an adjunct to anemia management has transformed the lives of patients with chronic renal failure on dialysis and reduced transfusion needs of anemic cancer patients receiving chemotherapy. Eventually, patients with inherited disorders of globin synthesis or mutations in the globin gene, such as sickle cell disease, may benefit from the successful introduction of targeted genetic therapy (Chap. 470).

POLYCYTHEMIA

Polycythemia is defined as an increase in the hemoglobin above normal. This increase may be real or only apparent because of a decrease in plasma volume (spurious or relative polycythemia). The term *erythrocytosis* may be used interchangeably with polycythemia, but some draw a distinction between them: erythrocytosis implies documentation of increased red cell mass, whereas polycythemia refers to any increase in red cells. Often patients with polycythemia are detected through an incidental finding of elevated hemoglobin or hematocrit levels. Concern that the hemoglobin level may be abnormally high is usually triggered at 17 g/dL (170 g/L) for men and 15 g/dL (150 g/L) for women. Hematocrit levels $>50\%$ in men or $>45\%$ in women may be abnormal. Hematocrits $>60\%$ in men and $>55\%$ in women are almost invariably associated with an increased red cell mass. Given that the machine that quantitates red cell parameters actually measures hemoglobin concentrations and calculates hematocrits, hemoglobin levels may be a better index.

Features of the clinical history that are useful in the differential diagnosis include smoking, current living at high altitude, a history of diuretic use, congenital heart disease, sleep apnea, or chronic lung disease.

Patients with polycythemia may be asymptomatic or experience symptoms related to the increased red cell mass or the underlying disease process that leads to the increased red cell mass. The dominant symptoms from an increased red cell mass are related to hyperviscosity and thrombosis (both venous and arterial), because the blood viscosity increases logarithmically at hematocrits $>55\%$. Manifestations include neurologic symptoms such as vertigo, tinnitus, headache, and visual disturbances. Hypertension is often present. Patients with *polycythemia vera* may have aquagenic pruritus, symptoms related to

hepatosplenomegaly, easy bruising, epistaxis, or bleeding from the gastrointestinal tract. Peptic ulcer disease is common. Such patients also may present with digital ischemia, Budd-Chiari syndrome, or hepatic or splenic/mesenteric vein thrombosis. Patients with hypoxemia may develop cyanosis on minimal exertion or have headache, impaired mental acuity, and fatigue.

The physical examination usually reveals a ruddy complexion. Splenomegaly favors polycythemia vera as the diagnosis (**Chap. 103**). The presence of cyanosis or evidence of a right-to-left shunt suggests congenital heart disease presenting in the adult, particularly tetralogy of Fallot or Eisenmenger's syndrome (**Chap. 269**). Increased blood viscosity raises pulmonary artery pressure; hypoxemia can lead to increased pulmonary vascular resistance. Together, these factors can produce cor pulmonale.

Polycythemia can be spurious (related to a decrease in plasma volume; Gaisböck's syndrome), primary, or secondary in origin. The secondary causes are all mediated by EPO: either a physiologically adapted appropriate level based on tissue hypoxia (lung disease, high altitude, CO poisoning, high-affinity hemoglobinopathy) or an abnormal overproduction (renal cysts, renal artery stenosis, tumors with ectopic EPO production). A rare familial form of polycythemia is associated with normal EPO levels but hyperresponsive EPO receptors due to mutations.

APPROACH TO THE PATIENT

Polycythemia

As shown in **Fig. 63-18**, the first step is to document the presence of an increased red cell mass using the principle of isotope dilution by administering ^{51}Cr -labeled autologous red blood cells to the patient and sampling blood radioactivity over a 2-h period. If the red cell mass is normal (<36 mL/kg in men, <32 mL/kg in women), the patient has spurious or relative polycythemia. If the red cell mass is

increased (>36 mL/kg in men, >32 mL/kg in women), serum EPO levels should be measured. It must be acknowledged that measurement of red cell mass is a physiologic approach to distinguishing polycythemia, and because of the use of radionuclide-labeled red cells, it is rarely performed. It is more common to measure EPO levels in a person with an elevated hemoglobin level or hematocrit. If EPO levels are low or unmeasurable, the patient most likely has polycythemia vera. A mutation in JAK2 (Val617Phe), a key member of the cytokine intracellular signaling pathway, can be found in 90–95% of patients with polycythemia vera. Many of those without this particular JAK2 mutation have mutations in exon 12. If EPO levels are low, check for JAK2 mutation(s), and perform an abdominal ultrasound to assess spleen size. Tests that support the diagnosis of polycythemia vera include elevated white blood cell count, increased absolute basophil count, and thrombocytosis. In practice, many physicians order EPO levels and assessment for JAK2 mutations at the same time.

If serum EPO levels are elevated, one needs to distinguish whether the elevation is a physiologic response to hypoxia or related to autonomous EPO production. Patients with low arterial O_2 saturation (<92%) should be further evaluated for the presence of heart or lung disease, if they are not living at high altitude. Patients with normal O_2 saturation who are smokers may have elevated EPO levels because of CO displacement of O_2 . If carboxyhemoglobin (COHb) levels are high, the diagnosis is "smoker's polycythemia." Such patients should be urged to stop smoking. Those who cannot stop smoking require phlebotomy to control their polycythemia. Patients with normal O_2 saturation who do not smoke either have an abnormal hemoglobin that does not deliver O_2 to the tissues (evaluated by finding elevated O_2 -hemoglobin affinity) or have a source of EPO production that is not responding to the normal feedback inhibition. Further workup is dictated by the differential diagnosis of EPO-producing neoplasms. Hepatoma, uterine leiomyoma, and renal cancer or cysts are all detectable with abdominopelvic computed tomography scans. Cerebellar hemangiomas may produce EPO, but they present with localizing neurologic signs and symptoms rather than polycythemia-related symptoms.

FURTHER READING

- HILLMAN RS et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.
- McMULLIN MF et al: Guidelines for the diagnosis, investigation and management of polycythaemia/erythrocytosis. *Br J Haematol* 130:174, 2005.
- SANKARAN VG, WEISS MJ: Anemia: progress in molecular mechanisms and therapies. *Nat Med* 21:221, 2015.
- SPIVAK JL: How I manage polycythemia vera. *Blood* 134:341, 2019.

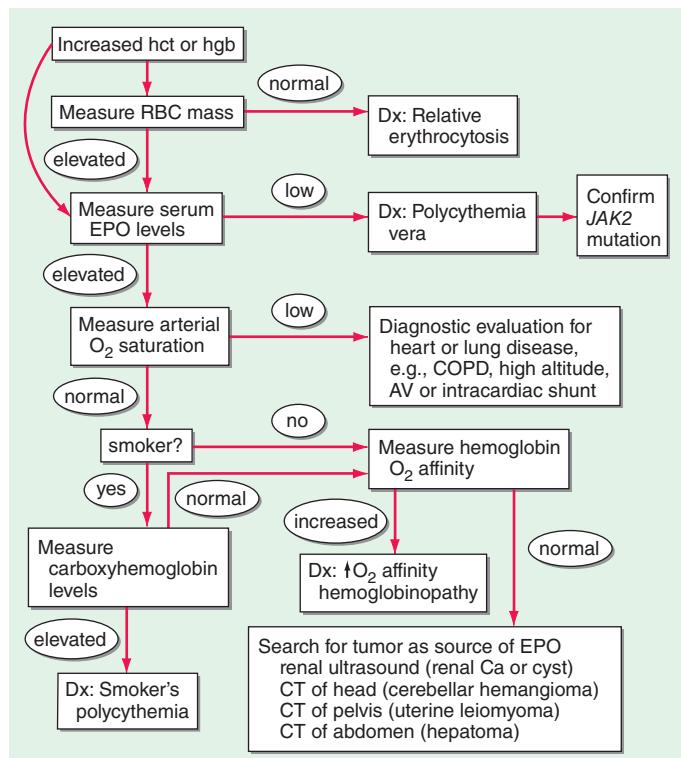


FIGURE 63-18 An approach to the differential diagnosis of patients with an elevated hemoglobin (possible polycythemia). AV, atrioventricular; Ca, calcium; COPD, chronic obstructive pulmonary disease; CT, computed tomography; EPO, erythropoietin; hct, hematocrit; hgb, hemoglobin; IVP, intravenous pyelogram; RBC, red blood cell.

64 Disorders of Granulocytes and Monocytes

Steven M. Holland, John I. Gallin

Leukocytes, the major cells comprising inflammatory and immune responses, include neutrophils, T and B lymphocytes, natural killer (NK) cells, monocytes, eosinophils, and basophils. These cells have specific functions, such as antibody production by B lymphocytes or destruction of bacteria by neutrophils, but in no single infectious disease is the exact role of the cell types completely established. Thus, whereas neutrophils are classically thought to be critical to host defense

against bacteria, they may also play important roles in defense against viral infections.

The blood delivers leukocytes to the various tissues from the bone marrow, where they are produced. Normal blood leukocyte counts are $4.3\text{--}10.8 \times 10^9/\text{L}$, with neutrophils representing 45–74% of the cells, bands 0–4%, lymphocytes 16–45%, monocytes 4–10%, eosinophils 0–7%, and basophils 0–2%. Variation among individuals and among different ethnic groups can be substantial, with lower leukocyte numbers for certain African-American ethnic groups. Lower granulocyte numbers in African-Americans are often in the 1500–2000/ μL range and are generally without sequelae. The condition is termed benign ethnic neutropenia. The lower number of granulocytes is associated with null expression of the Duffy antigen receptor for cytokines (DARC) gene, a receptor for malarial parasites, the absence of which conveys resistance to malaria. The various leukocytes are derived from a common stem cell in the bone marrow. Three-fourths of the nucleated cells of bone marrow are committed to the production of leukocytes. Leukocyte maturation in the marrow is under the regulatory control of a number of different factors, known as colony-stimulating factors (CSFs) and interleukins (ILs). Because an alteration in the number and type of leukocytes is often associated with disease processes, total white blood cell (WBC) count (cells per μL) and differential counts are informative. This chapter focuses on neutrophils, monocytes, and eosinophils. **Lymphocytes and basophils are discussed in Chaps. 349 and 353, respectively.**

NEUTROPHILS

MATURATION

Important events in neutrophil life are summarized in Fig. 64-1. In normal humans, neutrophils are produced only in the bone marrow. The minimum number of stem cells necessary to support hematopoiesis is estimated to be 400–500 at any one time. Human blood monocytes, tissue macrophages, and stromal cells produce CSFs, hormones required for the growth of monocytes and neutrophils in the bone marrow. The hematopoietic system not only produces enough neutrophils ($\sim 1.3 \times 10^{11}$ cells per 80-kg person per day) to carry out physiologic functions but also has a large reserve stored in the marrow, which can be mobilized in response to inflammation or infection. An increase in the number of blood neutrophils is called *neutrophilia*, and the presence of immature cells is termed a *shift to the left*. A decrease in the number of blood neutrophils is called *neutropenia*.

Neutrophils and monocytes evolve from pluripotent stem cells under the influence of cytokines and CSFs (Fig. 64-2). The proliferation phase through the metamyelocyte takes about 1 week, while the maturation phase from metamyelocyte to mature neutrophil takes another week. The myeloblast is the first recognizable precursor cell and is followed by the *promyelocyte*. The promyelocyte evolves when the classic lysosomal granules, called the *primary*, or *azurophil*, *granules* are produced. The primary granules contain hydrolases, elastase, myeloperoxidase, cathepsin G, cationic proteins, and bactericidal/permeability-increasing protein, which is important for killing gram-negative bacteria. Azurophil granules also contain *defensins*, a family of cysteine-rich polypeptides with broad antimicrobial activity against bacteria, fungi and certain enveloped viruses. The promyelocyte divides to produce the *myelocyte*, a cell responsible for the synthesis of the *specific*, or *secondary*, *granules*, which contain unique (specific) constituents such as lactoferrin, vitamin B₁₂-binding protein, membrane components of the reduced nicotinamide-adenine dinucleotide phosphate (NADPH) oxidase required for hydrogen peroxide production, histaminase, and receptors for certain chemoattractants and adherence-promoting factors (CR3) as well as receptors for the basement membrane component, laminin. The secondary granules do not contain acid hydrolases and therefore are not classic lysosomes. Packaging of secondary granule contents during myelopoiesis is controlled by CCAAT/enhancer binding protein- ϵ . Secondary granule contents are readily released extracellularly, and their mobilization is important in modulating inflammation. During the final stages of maturation, no cell division occurs, and the cell passes through the metamyelocyte stage and then to the band neutrophil with a sausage-shaped nucleus (Fig. 64-3). As the band cell matures, the nucleus assumes a lobulated configuration. The nucleus of neutrophils normally contains up to four segments (Fig. 64-4). Excessive segmentation (>5 nuclear lobes) may be a manifestation of folate or vitamin B₁₂ deficiency or the congenital neutropenia syndrome of warts, hypogammaglobulinemia, infections, and myelokathexis (WHIM) described below. The Pelger-Hüet anomaly (Fig. 64-5), an infrequent dominant benign inherited trait caused by heterozygous mutations in the lamin B receptor, results in neutrophils with distinctive bilobed nuclei that must be distinguished from band forms. Acquired bilobed nuclei, pseudo-Pelger-Hüet anomaly, can occur with acute infections or in myelodysplastic syndromes. The physiologic role of the normal multilobed nucleus of neutrophils is

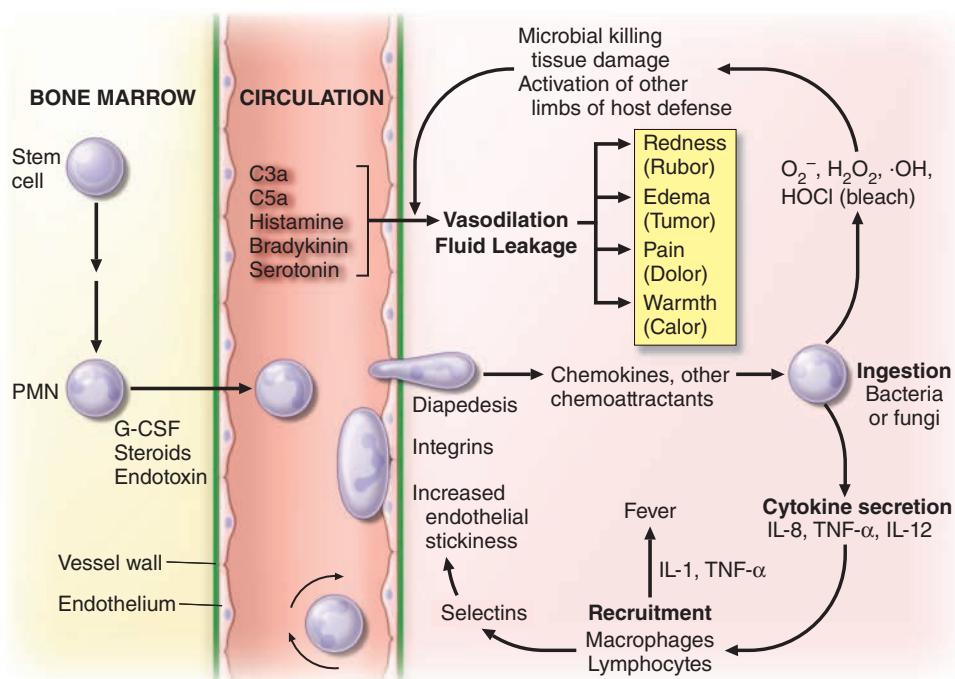


FIGURE 64-1 Schematic events in neutrophil production, recruitment, and inflammation. The four cardinal signs of inflammation (rubor, tumor, calor, dolor) are indicated, as are the interactions of neutrophils with other cells and cytokines. G-CSF, granulocyte colony-stimulating factor; IL, interleukin; PMN, polymorphonuclear leukocyte; TNF- α , tumor necrosis factor α .

Cell	Stage	Surface Markers ^a	Characteristics
	MYELOBLAST	CD33, CD13, CD15	Prominent nucleoli
	PROMYELOCYTE	CD33, CD13, CD15	Large cell Primary granules appear
	MYELOCYTE	CD33, CD13, CD15, CD14, CD11b	Secondary granules appear
	METAMYELOCYTE	CD33, CD13, CD15, CD14, CD11b	Kidney bean-shaped nucleus
	BAND FORM	CD33, CD13, CD15, CD14, CD11b, CD10, CD16	Condensed, band-shaped nucleus
	NEUTROPHIL	CD33, CD13, CD15, CD14, CD11b, CD10, CD16	Condensed, multilobed nucleus

^aCD = Cluster Determinant; ● Nucleolus; ● Primary granule; ● Secondary granule.

FIGURE 64-2 Stages of neutrophil development shown schematically. Granulocyte colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF) are critical to this process. Identifying cellular characteristics and specific cell-surface markers are listed for each maturational stage.

unknown, but it may allow great deformation of neutrophils during migration into tissues at sites of inflammation.

In severe acute bacterial infection, prominent neutrophil cytoplasmic granules, called *toxic granulations*, are occasionally seen.



FIGURE 64-3 Neutrophil band with Döhle body. The neutrophil with a sausage-shaped nucleus in the center of the field is a band form. Döhle bodies are discrete, blue-staining, nongranular areas found in the periphery of the cytoplasm of the neutrophil in infections and other toxic states. They represent aggregates of rough endoplasmic reticulum.

Toxic granulations are immature or abnormally staining azurophil granules. Cytoplasmic inclusions, also called *Döhle bodies* (Fig. 64-3), can be seen during infection and are fragments of ribosome-rich endoplasmic reticulum. Large neutrophil vacuoles are often present in acute bacterial infection in some viral infections such as COVID-19 and probably represent pinocytosed (internalized) membrane (Fig. 64-6).

Neutrophils are heterogeneous in function. Monoclonal antibodies have been developed that recognize only a subset of mature neutrophils. The meaning of neutrophil heterogeneity is not known.

The morphology of eosinophils and basophils is shown in Fig. 64-7.

MARROW RELEASE AND CIRCULATING COMPARTMENTS

Specific signals, including IL-1, tumor necrosis factor α (TNF- α), the CSFs, complement fragments, and chemokines, mobilize leukocytes from the bone marrow and deliver them to the blood in an unstimulated state. Under normal conditions, ~90% of the neutrophil pool is in the bone marrow, 2–3% in the circulation, and the remainder in the tissues (Fig. 64-8).

The circulating pool exists in two dynamic compartments: one freely flowing and one margined. The freely flowing pool is about one-half the neutrophils in the basal state and is composed of those cells that are in the blood and not in contact with the endothelium. Margined leukocytes are those that are in close physical contact with the endothelium (Fig. 64-9). In the pulmonary circulation, where an extensive capillary bed (~1000 capillaries per alveolus) exists, margination occurs because the capillaries are about the same size as a mature neutrophil. Therefore, neutrophil fluidity and deformability are necessary to make the transit through the pulmonary bed. Increased neutrophil rigidity and

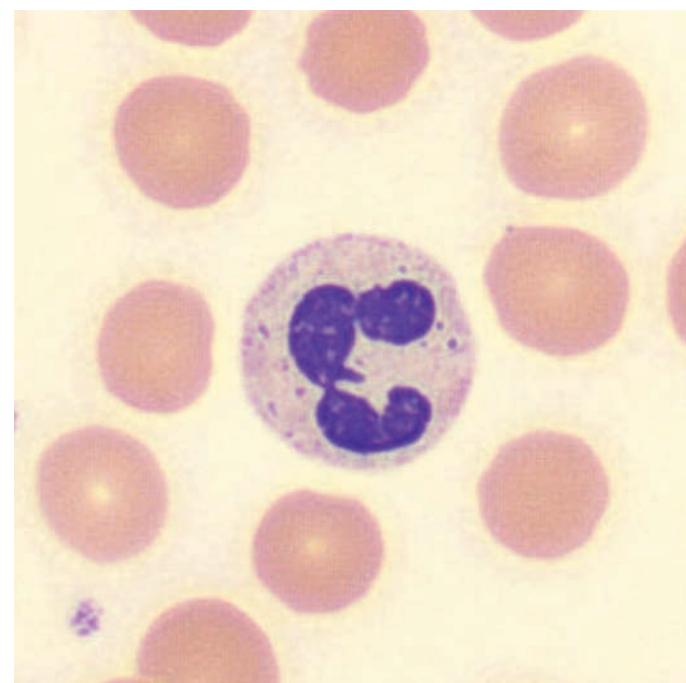


FIGURE 64-4 Normal granulocyte. The normal granulocyte has a segmented nucleus with heavy, clumped chromatin; fine neutrophilic granules are dispersed throughout the cytoplasm.

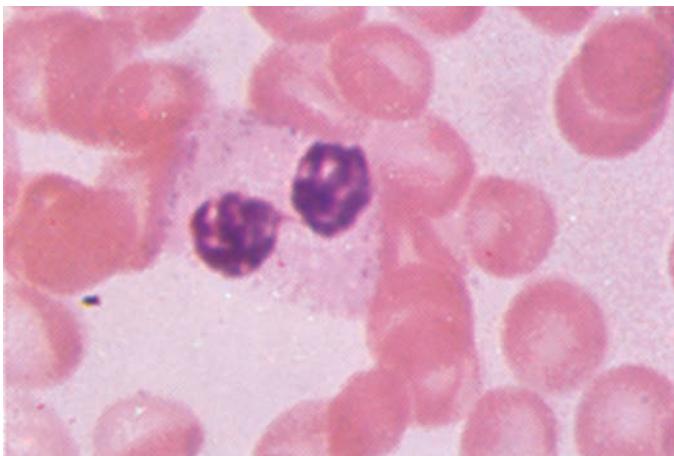


FIGURE 64-5 Pelger-Hüet anomaly. In this benign disorder, the majority of granulocytes are bilobed. The nucleus frequently has a spectacle-like, or “pince-nez,” configuration. (From M Lichtman et al (eds): *Williams Hematology*, 7th ed. New York, McGraw Hill, 2005; RS Hillman, KA Ault: *Hematology in General Practice*, 4th ed. New York, McGraw Hill, 2005.)

decreased deformability lead to augmented neutrophil trapping and margination in the lung. In contrast, in the systemic postcapillary venules, margination is mediated by the interaction of specific cell-surface molecules called *selectins*. Selectins are glycoproteins expressed on neutrophils and endothelial cells, among others, that cause a low-affinity interaction, resulting in “rolling” of the neutrophil along the endothelial surface. On neutrophils, the molecule L-selectin (cluster determinant [CD] 62L) binds to glycosylated proteins on endothelial cells (e.g., glycosylation-dependent cell adhesion molecule [GlyCAM1] and CD34). Glycoproteins on neutrophils, most importantly sialyl-Lewis^x (SL^x, CD15s), are targets for binding of selectins expressed on endothelial cells (E-selectin [CD62E] and P-selectin [CD62P]) and other leukocytes. In response to chemotactic stimuli from injured tissues (e.g., complement product C5a, leukotriene B₄, IL-8) or bacterial products (e.g., N-formylmethionylleucylphenylalanine [f-met-leu-phe]), neutrophil adhesiveness increases through mobilization of intracellular adhesion proteins stored in specific granules to the cell surface, and the cells “stick” to the endothelium through *integrins*. The integrins are leukocyte glycoproteins that exist as complexes of a common CD18 β chain with CD11a (LFA-1), CD11b (called Mac-1, CR3, or the C3bi receptor), and CD11c (called p150,95 or CR4). CD11a/CD18 and CD11b/CD18 bind to specific endothelial receptors (intercellular adhesion molecules [ICAM] 1 and 2).

On cell stimulation, L-selectin is shed from neutrophils, and E-selectin increases in the blood, presumably because it is shed from endothelial cells; receptors for chemoattractants and opsonins are mobilized; and the phagocytes orient toward the chemoattractant source in the extravascular space, increase their motile activity (chemokinesis),

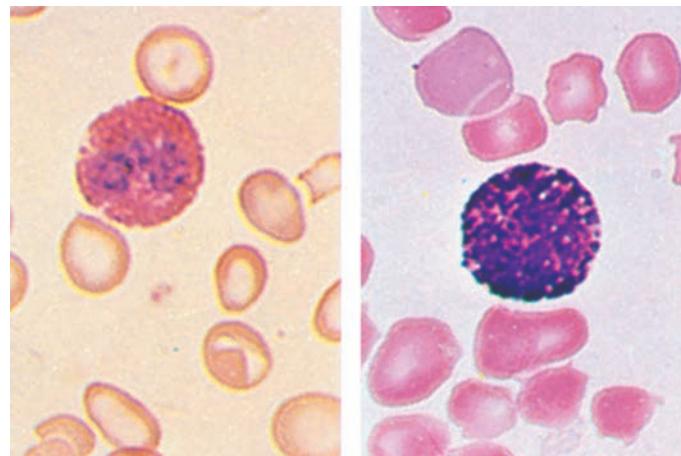


FIGURE 64-7 Normal eosinophil (left) and basophil (right). The eosinophil contains large, bright orange granules and usually a bilobed nucleus. The basophil contains large purple-black granules that fill the cell and obscure the nucleus.

and migrate directionally (chemotaxis) into tissues. The process of migration into tissues is called *diapedesis* and involves the crawling of neutrophils between postcapillary endothelial cells that open junctions between adjacent cells to permit leukocyte passage. Diapedesis involves platelet/endothelial cell adhesion molecule (PECAM) 1 (CD31), which is expressed on both the emigrating leukocyte and the endothelial cells. The endothelial responses (increased blood flow from increased vasodilation and permeability) are mediated by anaphylatoxins (e.g., C3a and C5a) as well as vasodilators such as histamine, bradykinin, serotonin, nitric oxide, vascular endothelial growth factor (VEGF), and prostaglandins E and I. Cytokines regulate some of these processes (e.g., TNF- α induction of VEGF, interferon [IFN] γ inhibition of prostaglandin E).

In the healthy adult, most neutrophils leave the body by migration through the mucous membrane of the gastrointestinal tract. Normally, neutrophils spend a short time in the circulation (half-life, 6–7 h). Senescent neutrophils are cleared from the circulation by macrophages in the lung and spleen. Once in the tissues, neutrophils release enzymes, such as collagenase and elastase, which may help establish abscess cavities. Neutrophils ingest pathogenic materials that have been opsonized by IgG and C3b. Fibronectin and the tetrapeptide tuftsin also facilitate phagocytosis.

With phagocytosis comes a burst of oxygen consumption and activation of the hexose-monophosphate shunt. A membrane-associated NADPH oxidase, consisting of membrane and cytosolic components, is assembled and catalyzes the univalent reduction of oxygen to superoxide anion, which is then converted by superoxide dismutase to hydrogen peroxide and other toxic oxygen products (e.g.,

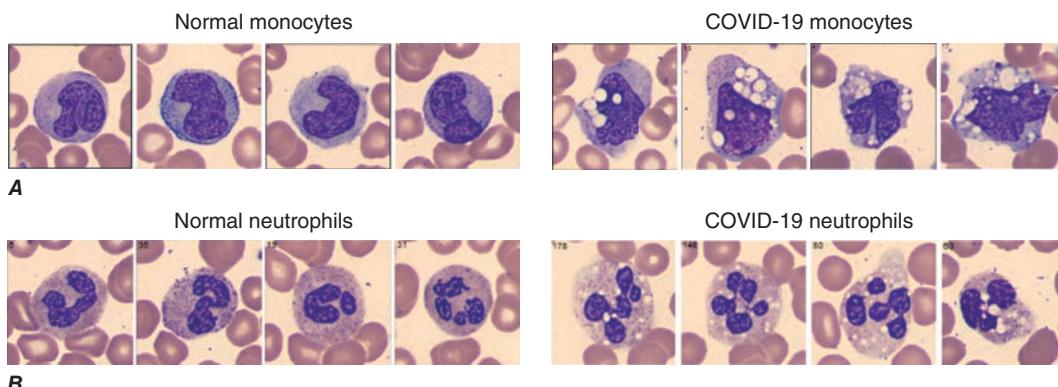


FIGURE 64-6 COVID-19: Vacuolization in peripheral blood monocytes and neutrophils of COVID-19 patients. Peripheral blood smear showing vacuolization in (A) monocytes and (B) neutrophils from hospitalized hypoxic COVID-19 patients relative to healthy volunteers. Increased vacuoles were noted in ~80% of monocytes and ~50% of neutrophils in each COVID-19 patient throughout their hospitalization.

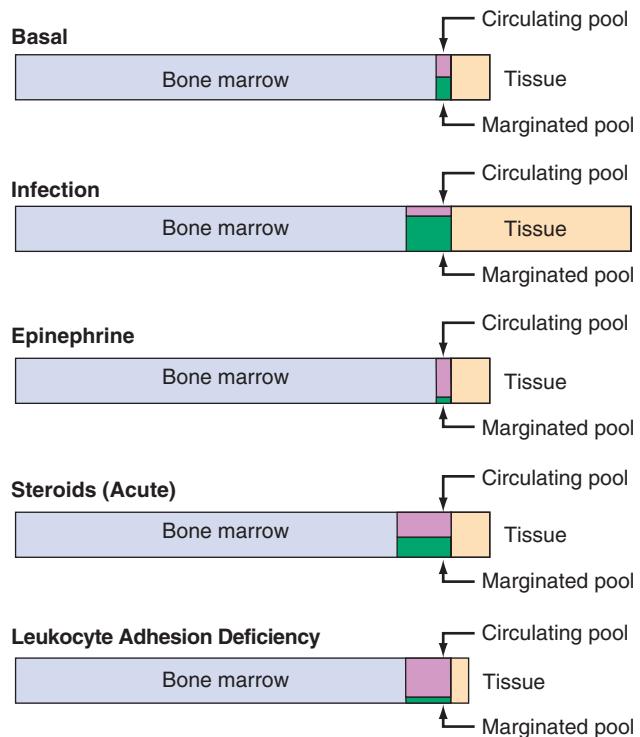


FIGURE 64-8 Schematic neutrophil distribution and kinetics between the different anatomic and functional pools.

hydroxyl radical). Hydrogen peroxide + chloride + neutrophil myeloperoxidase generates hypochlorous acid (bleach), hypochlorite, and chlorine. These products oxidize and halogenate microorganisms and tumor cells and, when uncontrolled, can damage host tissue. Strongly

cationic proteins, defensins, elastase, cathepsins, and probably nitric oxide also participate in microbial killing. Lactoferrin chelates iron, an important growth factor for microorganisms, especially fungi. Other enzymes, such as lysozyme and acid proteases, help digest microbial debris. After 1–4 days in tissues, neutrophils die. The apoptosis of neutrophils is also cytokine-regulated; granulocyte colony-stimulating factor (G-CSF) and IFN- γ prolong their life span. Neutrophil extracellular traps (NETs) consisting of a DNA scaffold decorated with neutrophil-granule derived proteins, such as enzymatically active proteases and antimicrobial peptides, have been described recently and are thought to be formed as a defense mechanism to immobilize invading microorganisms. Under certain conditions, such as in delayed-type hypersensitivity, monocyte accumulation occurs within 6–12 h of initiation of inflammation. Neutrophils, monocytes, microorganisms in various states of digestion, and altered local tissue cells make up the inflammatory exudate, pus. Myeloperoxidase confers the characteristic green color to pus and may participate in turning off the inflammatory process by inactivating chemoattractants and immobilizing phagocytic cells.

Neutrophils respond to certain cytokines (IFN- γ , granulocyte-macrophage colony-stimulating factor [GM-CSF], IL-8) and produce cytokines and chemotactic signals (TNF- α , IL-8, macrophage inflammatory protein [MIP] 1) that modulate the inflammatory response. In the presence of fibrinogen, f-met-leu-phe or leukotriene B₄ IL-8 production by neutrophils is induced, providing autocrine amplification of inflammation. *Chemokines (chemoattractant cytokines)* are small proteins produced by many different cell types, including endothelial cells, fibroblasts, epithelial cells, neutrophils, and monocytes, that regulate neutrophil, monocyte, eosinophil, and lymphocyte recruitment and activation. Chemokines transduce their signals through heterotrimeric G protein-linked receptors that have seven cell membrane-spanning domains, the same type of cell-surface receptor that mediates the response to the classic chemoattractants f-met-leu-phe and C5a. Four major groups of chemokines are recognized based on the cysteine structure near the N terminus: C, CC, CXC, and CXXXC. The CXC cytokines such as IL-8 mainly attract neutrophils; CC chemokines such as MIP-1 attract lymphocytes, monocytes, eosinophils, and basophils; the C chemokine lymphotactin is T-cell tropic; the CXXXC chemokine fractalkine attracts neutrophils, monocytes, and T cells. These molecules and their receptors not only regulate the trafficking and activation of inflammatory cells, but specific chemokine receptors also serve as co-receptors for HIV infection (Chap. 202), while others have roles in other viral infections (e.g., West Nile virus), susceptibility and response to *Candida*, and atherogenesis.

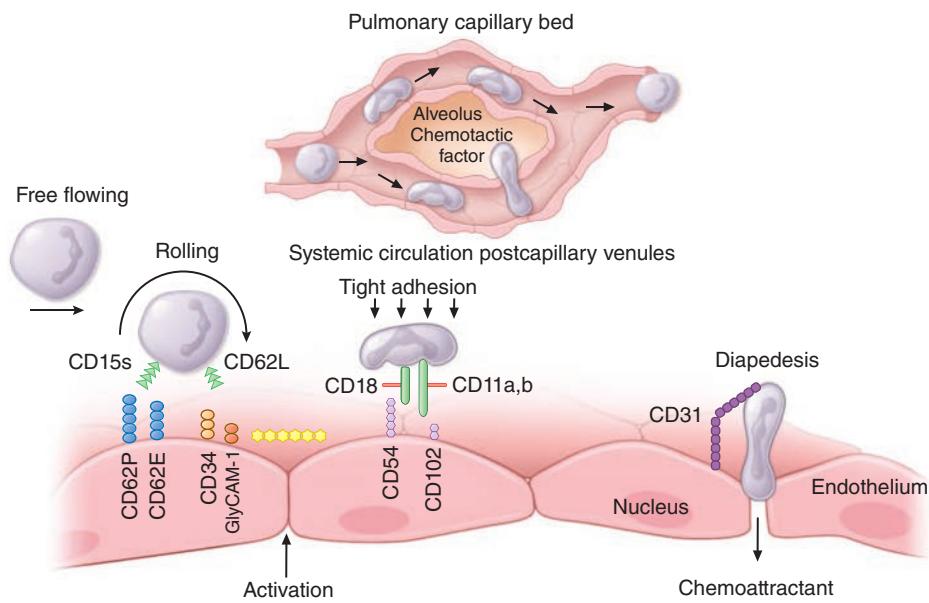


FIGURE 64-9 Neutrophil travel through the pulmonary capillaries is dependent on neutrophil deformability. Neutrophil rigidity (e.g., caused by C5a) enhances pulmonary trapping and response to pulmonary pathogens in a way that is not so dependent on cell-surface receptors. Intraalveolar chemotactic factors, such as those caused by certain bacteria (e.g., *Streptococcus pneumoniae*), lead to diapedesis of neutrophils from the pulmonary capillaries into the alveolar space. Neutrophil interaction with the endothelium of the systemic postcapillary venules is dependent on molecules of attachment. The neutrophil "rolls" along the endothelium using selectins: neutrophil CD15s (sialyl-Lewis^x) binds to CD62E (E-selectin) and CD62P (P-selectin) on endothelial cells; CD62L (L-selectin) on neutrophils binds to CD34 and other molecules (e.g., GlyCAM-1) expressed on endothelium. Chemokines or other activation factors stimulate integrin-mediated "tight adhesion": CD11a/CD18 (LFA-1) and CD11b/CD18 (Mac-1, CR3) bind to CD54 (ICAM-1) and CD102 (ICAM-2) on the endothelium. Diapedesis occurs between endothelial cells: CD31 (PECAM-1) expressed by the emigrating neutrophil interacts with CD31 expressed at the endothelial cell-cell junction. CD, cluster determinant; GlyCAM, glycosylation-dependent cell adhesion molecule; ICAM, intercellular adhesion molecule; PECAM, platelet/endothelial cell adhesion molecule.

■ NEUTROPHIL ABNORMALITIES

Defects in the neutrophil life cycle can lead to dysfunction and compromised host defenses. When inflammation is severely depressed the clinical result is often recurrent, severe bacterial and fungal infections. Aphthous ulcers of mucous membranes (gray ulcers without pus) and gingivitis and periodontal disease suggest a phagocytic cell disorder. Patients with congenital phagocyte defects can have infections within the first few days of life. Skin, ear, upper and lower respiratory tract, and bone infections are common. Sepsis and meningitis are rare. In some disorders, the frequency

of infection is variable, and patients can go for months or even years without major infection. Aggressive management of these congenital diseases, including hematopoietic stem cell transplantation and gene therapy, has extended the life span of patients well into adulthood.

Neutropenia The consequences of absent neutrophils are dramatic. Susceptibility to infectious diseases increases sharply when neutrophil counts fall to <1000 cells/ μ L. When the absolute neutrophil count (ANC; band forms and mature neutrophils combined) falls to <500 cells/ μ L, control of endogenous microbial flora (e.g., mouth, gut) is impaired; when the ANC is <200/ μ L, the local inflammatory process is absent. Neutropenia can be due to depressed production, increased peripheral destruction, or excessive peripheral pooling. A falling neutrophil count or a significant decrease in the number of neutrophils below steady-state levels, together with a failure to increase neutrophil counts in the setting of infection or other challenge, requires investigation. Acute neutropenia, such as that caused by cancer chemotherapy, is more likely to be associated with increased risk of infection than chronic neutropenia (months to years) that reverses in response to infection or carefully controlled administration of endotoxin (see “Laboratory Diagnosis and Management,” below).

Some causes of inherited and acquired neutropenia are listed in **Table 64-1**. The most common neutropenias are iatrogenic, resulting from the use of cytotoxic or immunosuppressive therapies for malignancy or control of autoimmune disorders. These drugs cause neutropenia because they result in decreased production of rapidly growing progenitor (stem) cells of the marrow. Certain antibiotics such as chloramphenicol, trimethoprim-sulfamethoxazole, flucytosine, vidarabine, and the antiretroviral drug zidovudine may cause neutropenia by inhibiting proliferation of myeloid precursors. Azathioprine and 6-mercaptopurine are metabolized by the enzyme thiopurine methyltransferase (TMPT); hypofunctional polymorphisms that are found in 11% of whites can lead to accumulation of 6-thioguanine and profound marrow toxicity. The marrow suppression is generally dose-related and dependent on continued administration of the drug. Cessation of the offending agent and recombinant human G-CSF usually reverse these forms of neutropenia.

Another important mechanism for iatrogenic neutropenia is the effect of drugs that serve as immune haptens and sensitize neutrophils or neutrophil precursors to immune-mediated peripheral

destruction. This form of drug-induced neutropenia can be seen within 7 days of exposure to the drug; with previous drug exposure, resulting in preexisting antibodies, neutropenia may occur a few hours after administration of the drug. Although any drug can cause this form of neutropenia, the most frequent causes are commonly used antibiotics, such as sulfa-containing compounds, penicillins, and cephalosporins. Fever and eosinophilia may also be associated with drug reactions, but often these signs are not present. Drug-induced neutropenia can be severe, but discontinuation of the sensitizing drug is sufficient for recovery, which is usually seen within 5–7 days and is complete by 10 days. Readministration of the sensitizing drug should be avoided, because abrupt neutropenia will often result. For this reason, diagnostic challenge should be avoided.

Autoimmune neutropenias caused by circulating antineutrophil antibodies are another form of acquired neutropenia that results in increased destruction of neutrophils. Acquired neutropenia may also be seen with viral infections, including acute infection with HIV. Acquired neutropenia may be cyclic in nature, occurring at intervals of several weeks. Acquired cyclic or stable neutropenia may be associated with an expansion of large granular lymphocytes (LGLs), which may be T cells, NK cells, or NK-like cells. Patients with large granular lymphocytosis may have moderate blood and bone marrow lymphocytosis, neutropenia, polyclonal hypergammaglobulinemia, splenomegaly, rheumatoid arthritis, and absence of lymphadenopathy. Such patients may have a chronic and relatively stable course. Recurrent bacterial infections are frequent. Benign and malignant forms of this syndrome occur. In some patients, a spontaneous regression has occurred even after 11 years, suggesting an immunoregulatory defect as the basis for at least one form of the disorder. Glucocorticoids, cyclosporine, methotrexate, and monoclonals are commonly used to manage these cytopenias.

Heredity Neutropenias Hereditary neutropenias are rare and may manifest in early childhood as a profound constant neutropenia or agranulocytosis. Congenital forms of neutropenia include Kostmann's syndrome (neutrophil count <100/ μ L), which is often fatal and due to mutations in the antiapoptosis gene *HAX-1*; severe chronic neutropenia (neutrophil count of 300–1500/ μ L) due to mutations in neutrophil elastase (*ELANE*); hereditary cyclic neutropenia, or, more appropriately, cyclic hematopoiesis, also due to mutations in neutrophil elastase (*ELANE*); the cartilage-hair hypoplasia syndrome due to mutations in the mitochondrial RNA-processing endoribonuclease *RMRP*; Shwachman-Diamond syndrome associated with pancreatic insufficiency due to mutations in the Shwachman-Bodian-Diamond syndrome gene *SBDS*; the WHIM (warts, hypogammaglobulinemia, infections, myelokathexis [retention of WBCs in the marrow]) syndrome, characterized by neutrophil hypersegmentation and bone marrow myeloid arrest due to mutations in the chemokine receptor *CXCR4*; and neutropenias associated with other immune defects, such as GATA2 deficiency, X-linked agammaglobulinemia, Wiskott-Aldrich syndrome, and CD40 ligand deficiency. Mutations in the G-CSF receptor can develop in severe congenital neutropenia and are linked to the development of leukemia. Absence of both myeloid and lymphoid cells is seen in reticular dysgenesis, due to mutations in the nuclear genome-encoded mitochondrial enzyme adenylate kinase-2 (*AK2*).

Maternal factors can be associated with neutropenia in the newborn. Transplacental transfer of IgG directed against antigens on fetal neutrophils can result in peripheral destruction. Drugs (e.g., thiazides) ingested during pregnancy can cause neutropenia in the newborn by either depressed production or peripheral destruction.

In Felty's syndrome—the triad of rheumatoid arthritis, splenomegaly, and neutropenia (**Chap. 358**)—spleen-produced antibodies can shorten neutrophil life span, while large granular lymphocytes can attack marrow neutrophil precursors. Splenectomy may increase the neutrophil count in Felty's syndrome and lower serum neutrophil-binding IgG. Some Felty's syndrome patients also have autoantibodies to G-CSF, while others have increased numbers of LGLs. Splenomegaly with peripheral trapping and destruction of neutrophils is also seen in lysosomal storage diseases and commonly in portal hypertension.

TABLE 64-1 Causes of Neutropenia

Decreased Production

Drug-induced—alkylating agents (nitrogen mustard, busulfan, chlorambucil, cyclophosphamide); antimetabolites (methotrexate, 6-mercaptopurine, 5-flucytosine); noncytotoxic agents (antibiotics [chloramphenicol, penicillins, sulfonamides], phenothiazines, tranquilizers [meprobamate], anticonvulsants [carbamazepine], antipsychotics [clozapine], certain diuretics, anti-inflammatory agents, antithyroid drugs, many others)

Hematologic diseases—idiopathic, cyclic neutropenia, Chédiak-Higashi syndrome, aplastic anemia, infantile genetic disorders (see text)

Tumor invasion, myelofibrosis

Nutritional deficiency—vitamin B₁₂, folate (especially alcoholics)

Infection—tuberculosis, typhoid fever, brucellosis, tularemia, measles, infectious mononucleosis, malaria, viral hepatitis, leishmaniasis, AIDS

Peripheral Destruction

Antineutrophil antibodies and/or splenic or lung trapping

Autoimmune disorders—Felty's syndrome, rheumatoid arthritis, lupus erythematosus

Drugs as haptens—aminopyrine, α -methyl-dopa, phenylbutazone, mercurial diuretics, some phenothiazines

Granulomatosis with polyangiitis (Wegener's)

Peripheral Pooling (Transient Neutropenia)

Overwhelming bacterial infection (acute endotoxemia)

Hemodialysis

Cardiopulmonary bypass

TABLE 64-2 Causes of Neutrophilia**Increased Production**

Idiopathic

Drug-induced—glucocorticoids, G-CSF

Infection—bacterial, fungal, sometimes viral

Inflammation—thermal injury, tissue necrosis, myocardial and pulmonary infarction, hypersensitivity states, collagen vascular diseases

Myeloproliferative diseases—myelocytic leukemia, myeloid metaplasia, polycythemia vera

Increased Marrow Release

Glucocorticoids

Acute infection (endotoxin)

Inflammation—thermal injury

Decreased or Defective Margination

Drugs—epinephrine, glucocorticoids, nonsteroidal anti-inflammatory agents

Stress, excitement, vigorous exercise

Leukocyte adhesion deficiency type 1 (CD18); leukocyte adhesion deficiency type 2 (selectin ligand, CD15s); leukocyte adhesion deficiency type 3 (FERMT3)

Miscellaneous

Metabolic disorders—ketoacidosis, acute renal failure, eclampsia, acute poisoning

Drugs—lithium

Other—metastatic carcinoma, acute hemorrhage or hemolysis

Abbreviation: G-CSF, granulocyte colony-stimulating factor.

Neutrophilia Neutrophilia results from increased neutrophil production, increased marrow release, or defective margination (**Table 64-2**). The most important acute cause of neutrophilia is infection. Neutrophilia from acute infection represents both increased production and increased marrow release. Increased production is also associated with chronic inflammation and certain myeloproliferative diseases. Increased marrow release and mobilization of the marginated leukocyte pool are induced by glucocorticoids. Release of epinephrine, as with vigorous exercise, excitement, or stress, will demarginate neutrophils in the spleen and lungs and double the neutrophil count in minutes. Cigarette smoking can elevate neutrophil counts above the normal range. Leukocytosis with cell counts of 10,000–25,000/ μ L occurs in response to infection and other forms of acute inflammation and results from both release of the marginated pool and mobilization of marrow reserves. Persistent neutrophilia with cell counts of \geq 30,000–50,000/ μ L is called a *leukemoid reaction*, a term often used to distinguish this degree of neutrophilia from leukemia. In a leukemoid reaction, the circulating neutrophils are usually mature and not clonally derived.

Abnormal Neutrophil Function Inherited and acquired abnormalities of phagocyte function are listed in **Table 64-3**. The resulting diseases are best considered in terms of the functional defects of adherence, chemotaxis, and microbicidal activity. The distinguishing features of the important inherited disorders of phagocyte function are shown in **Table 64-4**.

DISORDERS OF ADHESION Three main types of leukocyte adhesion deficiency (LAD) have been described. All are autosomal recessive and result in the inability of neutrophils to exit the circulation to sites of infection, leading to leukocytosis and increased susceptibility to infection (Fig. 64-9). Patients with LAD 1 have mutations in CD18, the common component of the integrins LFA-1, Mac-1, and p150,95, leading to a defect in tight adhesion between neutrophils and the endothelium. The heterodimer formed by CD18/CD11b (Mac-1) is also the receptor for the complement-derived opsonin C3bi (CR3). The CD18 gene is located on distal chromosome 21q. The severity of the defect determines the severity of clinical disease. Complete lack of expression of the leukocyte integrins results in a severe phenotype in which inflammatory stimuli do not increase the expression of leukocyte integrins on neutrophils or activated T and B cells. Neutrophils (and monocytes) from patients with LAD 1 adhere poorly to endothelial cells and protein-coated surfaces and exhibit defective spreading, aggregation, and chemotaxis. The inability of neutrophils to exit the vasculature to the tissue deprives the tissue macrophage of its expected neutrophil ingestion, leading to macrophage production of IL-23, which induces T-cell production of IL-17, a potent proinflammatory cytokine. These processes conspire to drive inflammation in LAD 1. Patients with LAD 1 have recurrent bacterial infections involving the skin, oral and genital mucosa, and respiratory and intestinal tracts; persistent leukocytosis (resting neutrophil counts of 15,000–20,000/ μ L) because cells do not marginate; and, in severe cases, a history of delayed separation of the umbilical stump. Infections, especially of the skin, may become necrotic with progressively enlarging borders, slow healing, and development of dysplastic scars. The most common bacteria are *Staphylococcus aureus* and enteric gram-negative bacteria. LAD 2 is caused by an abnormality of fucosylation of SLe^x (CD15s), the ligand on neutrophils that interacts with selectins on endothelial cells and is responsible for neutrophil rolling along the endothelium. Infection susceptibility in LAD 2 appears to be less severe than in LAD 1. LAD 2 is also known as *congenital disorder of glycosylation IIc* (CDGIIc) due to mutation in a GDP-fucose transporter (SLC35C1). LAD 3 is characterized by infection susceptibility, leukocytosis, and petechial hemorrhage due to impaired integrin activation caused by mutations in the gene FERMT3.

DISORDERS OF NEUTROPHIL GRANULES The most common neutrophil defect is myeloperoxidase deficiency, a primary granule defect inherited as an autosomal recessive trait; the incidence is \sim 1 in 2000 persons. Isolated myeloperoxidase deficiency is not associated with

TABLE 64-3 Types of Granulocyte and Monocyte Disorders

FUNCTION	CAUSE OF INDICATED DYSFUNCTION		
	DRUG-INDUCED	ACQUIRED	INHERITED
Adherence-aggregation	Aspirin, colchicine, alcohol, glucocorticoids, ibuprofen, piroxicam	Neonatal state, hemodialysis	Leukocyte adhesion deficiency types 1, 2, and 3
Deformability		Leukemia, neonatal state, diabetes mellitus, immature neutrophils	
Chemokinesis-chemotaxis	Glucocorticoids (high dose), auranofin, colchicine (weak effect), phenylbutazone, naproxen, indomethacin, interleukin 2	Thermal injury, malignancy, malnutrition, periodontal disease, neonatal state, systemic lupus erythematosus, rheumatoid arthritis, diabetes mellitus, sepsis, influenza virus infection, herpes simplex virus infection, acrodermatitis enteropathica, AIDS	Chédiak-Higashi syndrome, neutrophil-specific granule deficiency, hyper IgE-recurrent infection (Job's) syndrome (in some patients), Down's syndrome, α -mannosidase deficiency, leukocyte adhesion deficiencies, Wiskott-Aldrich syndrome
Microbicidal activity	Colchicine, cyclophosphamide, glucocorticoids (high dose), TNF- α -blocking antibodies	Leukemia, aplastic anemia, certain neutropenias, tuftsin deficiency, thermal injury, sepsis, neonatal state, diabetes mellitus, malnutrition, AIDS	Chédiak-Higashi syndrome, neutrophil-specific granule deficiency, chronic granulomatous disease, defects in IFN γ /IL-12 axis

Abbreviations: IFN γ , interferon γ ; IL, interleukin; TNF- α , tumor necrosis factor alpha.

TABLE 64-4 Inherited Disorders of Phagocyte Function: Differential Features

CLINICAL MANIFESTATIONS	CELLULAR OR MOLECULAR DEFECTS	DIAGNOSIS
Chronic Granulomatous Diseases (70% X-Linked, 30% Autosomal Recessive)		
Severe infections of skin, ears, lungs, liver, and bone with catalase-positive microorganisms such as <i>Staphylococcus aureus</i> , <i>Burkholderia cepacia</i> complex, <i>Aspergillus</i> spp., <i>Chromobacterium violaceum</i> ; often hard to culture organism; excessive inflammation with granulomas, frequent lymph node suppuration; granulomas can obstruct GI or GU tracts; gingivitis, aphthous ulcers, seborrheic dermatitis	No respiratory burst due to the lack of one of five NADPH oxidase subunits in neutrophils, monocytes, and eosinophils	DHR or NBT test; no superoxide and H ₂ O ₂ production by neutrophils; immunoblot for NADPH oxidase components; genetic detection
Chédiak-Higashi Syndrome (Autosomal Recessive)		
Recurrent pyogenic infections, especially with <i>S. aureus</i> ; many patients get lymphoma-like illness during adolescence; periodontal disease; partial oculocutaneous albinism, nystagmus, progressive peripheral neuropathy, cognitive impairment in some patients	Reduced chemotaxis and phagolysosome fusion, increased respiratory burst activity, defective egress from marrow, abnormal skin window; defect in <i>CHS1</i>	Giant primary granules in neutrophils and other granule-bearing cells (Wright's stain); genetic detection
Specific Granule Deficiency (Autosomal Recessive and Dominant)		
Recurrent infections of skin, ears, and sinopulmonary tract; delayed wound healing; decreased inflammation; bleeding diathesis	Abnormal chemotaxis, impaired respiratory burst and bacterial killing, failure to upregulate chemotactic and adhesion receptors with stimulation, defect in transcription of granule proteins; defect in <i>CEBPE</i> or <i>SMARCD2</i>	Lack of secondary (specific) granules in neutrophils (Wright's stain), no neutrophil-specific granule contents (i.e., lactoferrin), no defensins, platelet α granule abnormality; genetic detection
Myeloperoxidase Deficiency (Autosomal Recessive)		
Clinically normal except in patients with underlying disease such as diabetes mellitus; then candidiasis or other fungal infections	No myeloperoxidase due to pre- and posttranslational defects in myeloperoxidase deficiency	No peroxidase in neutrophils; genetic detection
Leukocyte Adhesion Deficiency		
Type 1: Delayed separation of umbilical cord, sustained neutrophilia, recurrent infections of skin and mucosa, gingivitis, periodontal disease	Impaired phagocyte adherence, aggregation, spreading, chemotaxis, phagocytosis of C3b-coated particles; defective production of CD18 subunit common to leukocyte integrins	Reduced phagocyte surface expression of the CD18-containing integrins with monoclonal antibodies against LFA-1 (CD18/CD11a), Mac-1 or CR3 (CD18/CD11b), p150,95 (CD18/CD11c); genetic detection
Type 2: Cognitive impairment, short stature, Bombay (hh) blood phenotype, recurrent infections, neutrophilia	Impaired phagocyte rolling along endothelium; due to defects in fucose transporter	Reduced phagocyte surface expression of Sialyl-Lewis ^x , with monoclonal antibodies against CD15s; genetic detection
Type 3: Petechial hemorrhage, recurrent infections	Impaired signaling for integrin activation resulting in impaired adhesion due to mutation in <i>FERMT3</i>	Reduced signaling for adhesion through integrins; genetic detection
Phagocyte Activation Defects (X-Linked and Autosomal Recessive)		
NEMO deficiency: mild hypohidrotic ectodermal dysplasia; broad-based immune defect: pyogenic and encapsulated bacteria, viruses, <i>Pneumocystis</i> , mycobacteria; X-linked	Impaired phagocyte activation by IL-1, IL-18, TLR, CD40L, TNF-α leading to problems with inflammation and antibody production	Poor in vitro response to endotoxin; impaired NF-κB activation; genetic detection
IRAK4 and MyD88 deficiency: susceptibility to pyogenic bacteria such as staphylococci, streptococci, clostridia; resistant to <i>Candida</i> ; autosomal recessive	Impaired phagocyte activation by endotoxin through TLR and other pathways; TNF-α signaling preserved	Poor in vitro response to endotoxin; lack of NF-κB activation by endotoxin; genetic detection
Hyper IgE-Recurrent Infection Syndrome (Autosomal Dominant) (Job's Syndrome)		
Eczematoid or pruritic dermatitis, "cold" skin abscesses, recurrent pneumonias with <i>S. aureus</i> with bronchopleural fistulae and cyst formation, mild eosinophilia, mucocutaneous candidiasis, characteristic facies, restrictive lung disease, scoliosis, delayed primary dental deciduation	Reduced chemotaxis in some patients, reduced memory T and B cells; mutation in <i>STAT3</i>	Somatic and immune features involving lungs, skeleton, and immune system; serum IgE >2000 IU/ml; genetic testing
DOCK8 deficiency (autosomal recessive), severe eczema, atopic dermatitis, cutaneous abscesses, HSV, HPV, and molluscum infections, severe allergies, cancer	Impaired T-cell proliferation to mitogens; mutation in <i>DOCK8</i>	Severe allergies, viral infections, high IgE, eosinophilia, low IgM, progressive lymphopenia, genetic detection
Mycobacterial Susceptibility (Autosomal Dominant and Recessive Forms)		
Severe extrapulmonary or disseminated infections with bacille Calmette-Guérin (BCG), nontuberculous mycobacteria, salmonella, histoplasmosis, coccidioidomycosis, poor granuloma formation	Inability to kill intracellular organisms due to low IFN-γ production or response; mutations in IFN-γ receptors, IL-12 receptors, IL-12 p40, <i>STAT1</i> , <i>NEMO</i> , <i>ISG15</i> , <i>GATA2</i>	Abnormally low or very high levels of IFN-γ receptor 1; functional assays of cytokine production and response; genetic detection
GATA2 Deficiency (Autosomal Dominant)		
Persistent or disseminated warts, disseminated mycobacterial disease, low monocytes, NK cells, B cells; hypoplastic myelodysplasia, leukemia, cytogenetic abnormalities, pulmonary alveolar proteinosis	Impaired macrophage activity, cytopenias; mutations in <i>GATA2</i>	Profound circulating moncytopenia, NK and B-cell cytopenias; genetic detection

Abbreviations: C/EBPε, CCAAT/enhancer binding protein-ε; DHR, dihydrorhodamine (oxidation test); DOCK8, dedicator of cytokinesis 8; GI, gastrointestinal; GU, genitourinary; HPV, human papillomavirus; HSV, herpes simplex virus; IFN, interferon; IL, interleukin; IRAK4, IL-1 receptor-associated kinase 4; LFA-1, leukocyte function-associated antigen 1; MyD88, myeloid differentiation primary response gene 88; NADPH, nicotinamide-adenine dinucleotide phosphate; NBT, nitroblue tetrazolium (dye test); NEMO, NF-κB essential modulator; NF-κB, nuclear factor-κB; NK, natural killer; STAT1–3, signal transducer and activator of transcription 1–3; TLR, Toll-like receptor; TNF, tumor necrosis factor.

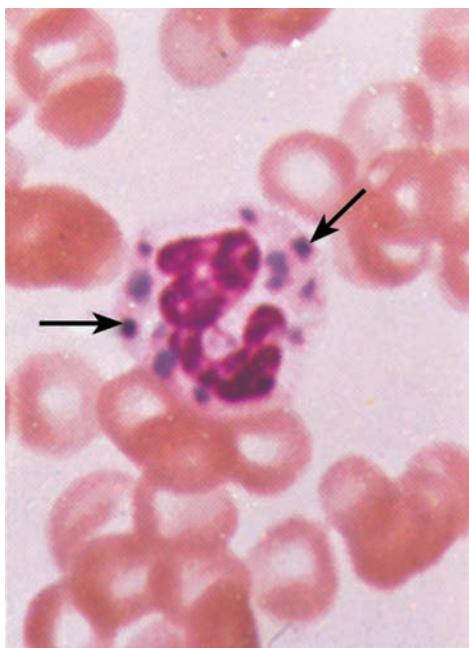


FIGURE 64-10 Chédiak-Higashi syndrome. The granulocytes contain huge cytoplasmic granules formed from aggregation and fusion of azurophilic and specific granules. Large abnormal granules are found in other granule-containing cells throughout the body.

clinically compromised defenses, presumably because other defense systems such as hydrogen peroxide generation are amplified. Microbicidal activity of neutrophils is delayed but not absent. Myeloperoxidase deficiency may make other acquired host defense defects more serious, and patients with myeloperoxidase deficiency and diabetes are more susceptible to *Candida* infections. An acquired form of myeloperoxidase deficiency occurs in myelomonocytic leukemia and acute myeloid leukemia.

Chédiak-Higashi syndrome (CHS) is a rare disease with autosomal recessive inheritance due to defects in the lysosomal transport protein LYST, encoded by the gene *CHS1* at 1q42. This protein is required for normal packaging and disbursement of granules. Neutrophils (and all cells containing lysosomes) from patients with CHS characteristically have large granules (Fig. 64-10), making it a systemic disease. Patients with CHS have nystagmus, partial oculocutaneous albinism, and an increased number of infections resulting from many bacterial agents. Some CHS patients develop an “accelerated phase” in childhood with a hemophagocytic syndrome and an aggressive lymphoma requiring bone marrow transplantation. CHS neutrophils and monocytes have impaired chemotaxis and abnormal rates of microbial killing due to slow rates of fusion of the lysosomal granules with phagosomes. NK cell function is also impaired. CHS patients may develop a severe disabling peripheral neuropathy in adulthood.

Specific granule deficiency is a rare autosomal recessive disease in which the production of secondary granules and their contents, as well as the primary granule component defensins, is defective. The defect in killing leads to severe bacterial infections. One type of specific granule deficiency is due to a mutation in the CCAAT/enhancer binding protein- ϵ , a regulator of expression of granule components. A dominant mutation in *C/EBP- ϵ* has also been described. Another form is caused by mutations in *SMARCD2*.

CHRONIC GRANULOMATOUS DISEASE Chronic granulomatous disease (CGD) is a group of disorders of granulocyte and monocyte oxidative metabolism due to a defect in the enzyme NADPH oxidase also called NOX2. Although CGD is rare, with an incidence of ~1 in 200,000 individuals, it is an important model of defective neutrophil oxidative metabolism. In about two-thirds of patients, CGD is inherited as an X-linked recessive trait; the remainder inherit their disease in autosomal recessive patterns. Mutations in the genes for the six proteins that allow assembly at the plasma membrane of NOX2 account

for all patients with CGD. Two proteins (a 91-kDa protein, abnormal in X-linked CGD, and a 22-kDa protein, absent in one form of autosomal recessive CGD) form the heterodimer cytochrome b-558 in the plasma membrane. The protein essential for reactive oxidant signaling (EROS) is encoded by *CYBC1*, which is required to transport the 91- and 22-kDa proteins to the endoplasmic reticulum. Three other proteins (40, 47, and 67 kDa, abnormal in the other autosomal recessive forms of CGD) are cytoplasmic in origin and interact with the cytochrome after cell activation to form the NADPH oxidase, required for hydrogen peroxide production. Leukocytes from patients with CGD have severely diminished hydrogen peroxide production. The genes involved in each of the defects have been cloned and sequenced and the chromosome locations identified. Patients with CGD characteristically have increased numbers of infections due to catalase-positive microorganisms (organisms that destroy their own hydrogen peroxide) such as *S. aureus*, *Serratia marsescens*, *Burkholderia cepacia* complex, *Nocardia* and *Aspergillus* species. When patients with CGD become infected, they often have extensive inflammatory reactions, and suppuration is common despite the administration of appropriate antibiotics. Aphthous ulcers and chronic inflammation of the nares are often present. Granulomas are frequent and can obstruct the gastrointestinal or genitourinary tracts. The excessive inflammation is due to failure to downregulate inflammation, reflecting a failure to inhibit the synthesis of, degradation of, or response to ILs or chemoattractants, leading to persistent myeloid reaction. Impaired killing of intracellular microorganisms by macrophages may lead to persistent cell-mediated immune activation and granuloma formation. Autoimmune complications such as immune thrombocytopenic purpura and juvenile idiopathic arthritis are also increased in CGD. In addition, for unexplained reasons, discoid lupus is more common in X-linked carriers. Late complications, including nodular regenerative hyperplasia and portal hypertension, are increasingly recognized in adolescent and adult patients with CGD. Interestingly, patients with CGD have been reported to be protected from atherosclerosis, suggesting an important role for NADPH oxidase (NOX2) in the pathogenesis of this inflammatory disease of arteries.

DISORDERS OF PHAGOCYTE ACTIVATION Phagocytes depend on cell-surface stimulation to induce signals that evoke multiple levels of the inflammatory response, including cytokine synthesis, chemotaxis, and antigen presentation. Mutations affecting the major pathway that signals through NF- κ B have been noted in patients with a variety of infection susceptibility syndromes. If the defects are at a very late stage of signal transduction, in the protein critical for NF- κ B activation known as the NF- κ B essential modulator (NEMO), then affected males develop ectodermal dysplasia and severe immune deficiency with susceptibility to bacteria, fungi, mycobacteria, and viruses. If the defects in NF- κ B activation are closer to the cell-surface receptors, in the proteins transducing Toll-like receptor signals, IL-1 receptor-associated kinase 4 (IRAK4), and myeloid differentiation primary response gene 88 (MyD88), then children have a marked susceptibility to pyogenic infections early in life but develop resistance to infection later.

MONONUCLEAR PHAGOCYTES

The mononuclear phagocyte system is composed of monoblasts, promonocytes, and monocytes, in addition to the structurally diverse tissue macrophages that make up what was previously referred to as the reticuloendothelial system. Macrophages are long-lived phagocytic cells capable of many of the functions of neutrophils. They are also secretory cells that participate in many immunologic and inflammatory processes distinct from neutrophils. Monocytes leave the circulation by diapedesis more slowly than neutrophils and have a half-life in the blood of 12–24 h.

Many tissue macrophages (“big eaters”) arise even before hematopoiesis and take up residence in tissues. In addition, there are macrophages derived from monocytes, which may have specialized functions suited for specific anatomic locations. Macrophages are particularly abundant in capillary walls of the lung, spleen, liver, and bone marrow, where they function to remove microorganisms and other noxious elements from the blood. Alveolar macrophages, liver Kupffer cells,

splenic macrophages, peritoneal macrophages, bone marrow macrophages, lymphatic macrophages, brain microglial cells, and dendritic macrophages all have specialized functions. Macrophage-secreted products include lysozyme, neutral proteases, acid hydrolases, arginase, complement components, enzyme inhibitors (plasmin, α_2 -macroglobulin), binding proteins (transferrin, fibronectin, transcobalamin II), nucleosides, and cytokines (TNF- α ; IL-1, 8, 12, 18). IL-1 (**Chaps. 18 and 349**) has many functions, including initiating fever in the hypothalamus, mobilizing leukocytes from the bone marrow, and activating lymphocytes and neutrophils. TNF- α is a pyrogen that duplicates many of the actions of IL-1 and plays an important role in the pathogenesis of gram-negative shock (**Chap. 304**). TNF- α stimulates production of hydrogen peroxide and related toxic oxygen species by macrophages and neutrophils. In addition, TNF- α induces catabolic changes that contribute to the profound wasting (cachexia) associated with many chronic diseases.

Other macrophage-secreted products include reactive oxygen and nitrogen metabolites, bioactive lipids (arachidonic acid metabolites and platelet-activating factors), chemokines, CSFs, and factors stimulating fibroblast and vessel proliferation. Macrophages help regulate the replication of lymphocytes and participate in the killing of tumors, viruses, and certain bacteria (*Mycobacterium tuberculosis* and *Listeria monocytogenes*). Macrophages are key effector cells in the elimination of intracellular microorganisms. Their ability to fuse to form giant cells that coalesce into granulomas in response to some inflammatory stimuli is important in the elimination of intracellular microbes and is under the control of IFN- γ . Nitric oxide induced by IFN- γ may be an important effector against intracellular parasites, including tuberculosis and *Leishmania*.

Macrophages play an important role in the immune response (**Chap. 349**). They process and present antigen to lymphocytes and secrete cytokines that modulate and direct lymphocyte development and function. Macrophages participate in autoimmune phenomena by removing immune complexes and other substances from the circulation. Polymorphisms in macrophage receptors for immunoglobulin (Fc γ RII) determine susceptibility to some infections and autoimmune diseases. In wound healing, they dispose of senescent cells, and they also contribute to atheroma development. Macrophage elastase mediates development of emphysema from cigarette smoking.

DISORDERS OF THE MONONUCLEAR PHAGOCYTE SYSTEM

Many disorders of neutrophils extend to mononuclear phagocytes. Monocytosis is associated with tuberculosis, brucellosis, subacute bacterial endocarditis, Rocky Mountain spotted fever, malaria, and visceral leishmaniasis (kala azar). Monocytosis also occurs with malignancies, leukemias, myeloproliferative syndromes, hemolytic anemias, chronic idiopathic neutropenias, and granulomatous diseases such as sarcoidosis, regional enteritis, and some collagen vascular diseases. Patients with LAD, hyperimmunoglobulin E-recurrent infection (Job's) syndrome, CHS, and CGD all have defects in the mononuclear phagocyte system.

Monocyte cytokine production or response is impaired in some patients with disseminated nontuberculous mycobacterial infection who are not infected with HIV. Genetic defects in the pathways regulated by IFN- γ and IL-12 lead to impaired killing of intracellular bacteria, mycobacteria, salmonellae, and certain viruses (**Fig. 64-11**).

Certain viral infections impair mononuclear phagocyte function. For example, influenza virus infection causes abnormal monocyte chemotaxis. Mononuclear phagocytes can be infected by HIV using CCR5, the chemokine receptor that acts as a co-receptor with CD4 for HIV. T lymphocytes produce IFN- γ , which induces FcR expression and phagocytosis and stimulates hydrogen peroxide production by mononuclear phagocytes and neutrophils. In certain diseases, such as AIDS, IFN- γ production may be deficient, whereas in other diseases, such as T-cell lymphomas, excessive release of IFN- γ may be associated with erythrophagocytosis by splenic macrophages.

Autoinflammatory diseases are characterized by abnormal cytokine regulation, leading to excess inflammation in the absence of infection. These diseases can mimic infectious or immunodeficient syndromes.

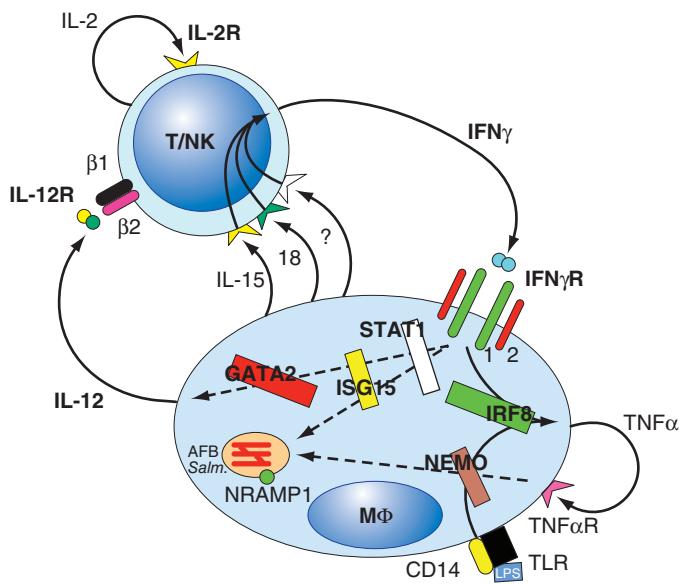


FIGURE 64-11 Lymphocyte-macrophage interactions underlying resistance to mycobacteria and other intracellular pathogens such as *Salmonella*, *Histoplasma*, and *Coccidioides*. Mycobacteria (and others) infect macrophages, leading to the production of IL-12, which activates T or NK cells through its receptor, leading to production of IL-2 and IFN- γ . IFN- γ acts through its receptor on macrophages to upregulate TNF- γ and IL-12 and kill intracellular pathogens. Other critical interacting molecules include signal transducer and activator of transcription 1 (STAT1), interferon regulatory factor 8 (IRF8), GATA2, and ISG15. Mutant forms of the cytokines and receptors shown in bold type have been found in severe cases of nontuberculous mycobacterial infection, salmonellosis, and other intracellular pathogens. AFB, acid-fast bacilli; IFN, interferon; IL, interleukin; NEMO, nuclear factor- κ B essential modulator; NK, natural killer; TLR, Toll-like receptor; TNF, tumor necrosis factor.

Gain-of-function mutations in the TNF- α receptor cause TNF- α receptor-associated periodic syndrome (TRAPS), which is characterized by recurrent fever in the absence of infection, due to persistent stimulation of the TNF- α receptor (**Chap. 369**). Diseases with abnormal IL-1 regulation leading to fever include familial Mediterranean fever due to mutations in PYRIN. Mutations in cold-induced autoinflammatory syndrome 1 (CIAS1) lead to neonatal-onset multisystem autoinflammatory disease, familial cold urticaria, and Muckle-Wells syndrome. The syndrome of pyoderma gangrenosum, acne, and sterile pyogenic arthritis (PAPA syndrome) is caused by mutations in PSTPIP1. In contrast to these syndromes of overexpression of proinflammatory cytokines, blockade of TNF- α by the antagonists infliximab, adalimumab, certolizumab, golimumab, or etanercept has been associated with severe infections due to tuberculosis, nontuberculous mycobacteria, and fungi (**Chap. 369**).

Monocytopenia occurs with acute infections, with stress, and after treatment with glucocorticoids. Drugs that suppress neutrophil production in the bone marrow can cause monocytopenia. Persistent severe circulating monocytopenia is seen in GATA2 deficiency, even though macrophages are found at the sites of inflammation. Monocytopenia also occurs in aplastic anemia, hairy cell leukemia, acute myeloid leukemia, and as a direct result of myelotoxic drugs.

EOSINOPHILS

Eosinophils and neutrophils share similar morphology, many lysosomal constituents, phagocytic capacity, and oxidative metabolism. Eosinophils express a specific chemoattractant receptor and respond to a specific chemokine, eotaxin, but little is known about their required role. Eosinophils are much longer lived than neutrophils, and unlike neutrophils, tissue eosinophils can recirculate. During most infections, eosinophils appear unimportant. However, in invasive helminthic infections, such as hookworm, schistosomiasis, strongyloidiasis, toxocariasis, trichinosis, filariasis, echinococcosis, and cysticercosis, the eosinophil plays a central role in host defense. Eosinophils are associated

with bronchial asthma, cutaneous allergic reactions, and other hypersensitivity states.

The distinctive feature of the red-staining (Wright's stain) eosinophil granule is its crystalline core consisting of an arginine-rich protein (major basic protein) with histaminase activity, important in host defense against parasites. Eosinophil granules also contain a unique eosinophil peroxidase that catalyzes the oxidation of many substances by hydrogen peroxide and may facilitate killing of microorganisms.

Eosinophil peroxidase, in the presence of hydrogen peroxide and halide, initiates mast cell secretion in vitro and thereby promotes inflammation. Eosinophils contain cationic proteins, some of which bind to heparin and reduce its anticoagulant activity. Eosinophil-derived neurotoxin and eosinophil cationic protein are ribonucleases that can kill respiratory syncytial virus. Eosinophil cytoplasm contains Charcot-Leyden crystal protein, a hexagonal bipyramidal crystal first observed in a patient with leukemia and then in sputum of patients with asthma; this protein is lysophospholipase and may function to detoxify certain lysophospholipids.

Several factors enhance the eosinophil's function in host defense. T cell-derived factors enhance the ability of eosinophils to kill parasites. Mast cell-derived eosinophil chemotactic factor of anaphylaxis (ECFa) increases the number of eosinophil complement receptors and enhances eosinophil killing of parasites. Eosinophil CSFs (e.g., IL-5) produced by macrophages increase eosinophil production in the bone marrow and activate eosinophils to kill parasites.

■ EOSINOPHILIA

Eosinophilia is the presence of >500 eosinophils per μL of blood and is common in many settings besides parasite infection. Significant tissue eosinophilia can occur without an elevated blood count. A common cause of eosinophilia is allergic reaction to drugs (iodides, aspirin, sulfonamides, nitrofurantoin, penicillins, and cephalosporins). Allergies such as hay fever, asthma, eczema, serum sickness, allergic vasculitis, and pemphigus are associated with eosinophilia. Eosinophilia also occurs in collagen vascular diseases (e.g., rheumatoid arthritis, eosinophilic fasciitis, allergic angiitis, and periarteritis nodosa) and malignancies (e.g., Hodgkin's disease; mycosis fungoidea; chronic myeloid leukemia; and cancer of the lung, stomach, pancreas, ovary, or uterus), as well as in STAT3-deficient Job's syndrome, DOCK8 deficiency (see below), and CGD. Eosinophilia is commonly present in helminthic infections. IL-5 is the dominant eosinophil growth factor. Therapeutic administration of the cytokines IL-2 or GM-CSF frequently leads to transient eosinophilia. The most dramatic hypereosinophilic syndromes are Loeffler's syndrome, tropical pulmonary eosinophilia, Loeffler's endocarditis, eosinophilic leukemia, and idiopathic hypereosinophilic syndrome (50,000–100,000/ μL). IL-5 is the dominant eosinophil growth factor and can be specifically inhibited with the monoclonal antibody mepolizumab.

The idiopathic hypereosinophilic syndromes are a heterogeneous group of disorders with the common feature of prolonged eosinophilia of unknown cause and organ system dysfunction, including the heart, central nervous system, kidneys, lungs, gastrointestinal tract, and skin. The bone marrow is involved in all affected individuals, but the most severe complications involve the heart and central nervous system. Clinical manifestations and organ dysfunction are highly variable. Eosinophils are found in the involved tissues and likely cause tissue damage by local deposition of toxic eosinophil proteins such as eosinophil cationic protein and major basic protein. In the heart, the pathologic changes lead to thrombosis, endocardial fibrosis, and restrictive endomyocardopathy. The damage to tissues in other organ systems is similar. Some cases are due to mutations involving the platelet-derived growth factor receptor, and these are extremely sensitive to the tyrosine kinase inhibitor imatinib. Glucocorticoids, hydroxyurea, and IFN- α each have been used successfully, as have therapeutic antibodies against IL-5. Cardiovascular complications are managed aggressively.

The *eosinophilia-myalgia syndrome* is a multisystem disease, with prominent cutaneous, hematologic, and visceral manifestations, that frequently evolves into a chronic course and can occasionally be fatal. The syndrome is characterized by eosinophilia (eosinophil count

>1000/ μL) and generalized disabling myalgias without other recognized causes. Eosinophilic fasciitis, pneumonitis, and myocarditis; neuropathy culminating in respiratory failure; and encephalopathy may occur. The disease is caused by ingesting contaminants in L-tryptophan-containing products. Eosinophils, lymphocytes, macrophages, and fibroblasts accumulate in the affected tissues, but their role in pathogenesis is unclear. Activation of eosinophils and fibroblasts and the deposition of eosinophil-derived toxic proteins in affected tissues may contribute. IL-5 and transforming growth factor β have been implicated as potential mediators. Treatment is withdrawal of products containing L-tryptophan and the administration of glucocorticoids. Most patients recover fully, remain stable, or show slow recovery, but the disease can be fatal in up to 5% of patients.

Eosinophilic neoplasms are discussed in Chap. 110.

■ EOSINOPENIA

Eosinopenia occurs with stress, such as acute bacterial infection, and after treatment with glucocorticoids. The mechanism of eosinopenia of acute bacterial infection is unknown but is independent of endogenous glucocorticoids, because it occurs in animals after total adrenalectomy. There is no known adverse effect of eosinopenia.

HYPERIMMUNOGLOBULIN E-RECURRENT INFECTION SYNDROME

The hyperimmunoglobulin E-recurrent infection syndrome, or Job's syndrome, is a rare multisystem disease in which the immune and somatic systems are affected, including neutrophils, monocytes, T cells, B cells, and osteoclasts. Autosomal *dominant* inhibitory mutations in signal transducer and activator of transcription 3 (STAT3) lead to inhibition of normal STAT signaling with broad and profound effects. Patients have characteristic facies with broad nose, kyphoscoliosis, and eczema. The primary teeth erupt normally but do not deciduate, often requiring extraction. Patients develop recurrent sinopulmonary and cutaneous infections that tend to be much less inflamed than appropriate for the degree of infection and have been referred to as "cold abscesses." Characteristically, pneumonias cavitate, leading to pneumatoceles. Coronary artery aneurysms are common, as are cerebral demyelinated plaques that accumulate with age. Importantly, IL-17-producing T cells, which are thought responsible for protection against extracellular and mucosal infections, are profoundly reduced in Job's syndrome. Despite very high IgE levels, these patients have only mildly elevated levels of allergy. An important syndrome with clinical overlap with the dominant negative STAT3 deficiency is due to autosomal recessive defects in dedicator of cytokinesis 8 (DOCK8). In DOCK8 deficiency, IgE elevation is joined to severe allergy, viral susceptibility, and increased rates of cancer. Autosomal dominant *gain-of-function* mutations in STAT3 lead to a disease characterized by onset in childhood of lymphadenopathy, autoimmune cytopenias, multiorgan autoimmunity, infections, and interstitial lung disease.

LABORATORY DIAGNOSIS AND MANAGEMENT

Initial studies of WBC and differential are essential, and careful examination of neutrophils on peripheral blood smears can diagnose Chediak-Higashi syndrome and suggest other neutrophil granule abnormalities such as specific granule deficiency. Often a bone marrow examination and serologies may be followed by either gene panel or whole exome sequencing in the cases of suspected genetic defects. Functionally, assessment of bone marrow reserves (steroid challenge test), marginated circulating pool of cells (epinephrine challenge test), and marginating ability (endotoxin challenge test) (Fig. 64-8) are also doable. In vivo assessment of inflammation is possible with a Rebuck skin window test or an in vivo skin blister assay, which measures the ability of leukocytes and inflammatory mediators to accumulate locally in the skin. In vitro tests of phagocyte aggregation, adherence, chemotaxis, phagocytosis, degranulation, and microbial activity (for *S. aureus*) may help pinpoint cellular or humoral lesions. Deficiencies of oxidative metabolism are detected with either the nitroblue tetrazolium (NBT) dye test or the dihydrorhodamine (DHR) oxidation test. These

tests are based on the ability of products of oxidative metabolism to alter the oxidation states of reporter molecules so that they can be detected microscopically (NBT) or by flow cytometry (DHR). Qualitative studies of superoxide and hydrogen peroxide production may further define neutrophil oxidative function.

Patients with leukopenias or leukocyte dysfunction often have delayed inflammatory responses. Therefore, clinical manifestations may be minimal despite overwhelming infection, and unusual infections must always be suspected. Early signs of infection demand prompt, aggressive culturing for microorganisms, use of antibiotics, and drainage of abscesses. Prolonged courses of antibiotics are often required. In patients with CGD, prophylactic antibiotics (trimethoprim-sulfamethoxazole) and antifungals (itraconazole) markedly diminish the frequency of life-threatening infections. Glucocorticoids may relieve gastrointestinal or genitourinary tract obstruction by granulomas in patients with CGD. Although TNF- α -blocking agents may markedly relieve inflammatory bowel symptoms, extreme caution must be exercised in their use in CGD inflammatory bowel disease, because it profoundly increases these patients' already heightened susceptibility to infection. Recombinant human IFN- γ , which nonspecifically stimulates phagocytic cell function, reduces the frequency of infections in patients with CGD by 70% and reduces the severity of infection. This effect of IFN- γ in CGD is additive to the effect of prophylactic antibiotics. The recommended dose is 50 μ g/m 2 subcutaneously three times weekly. IFN- γ has also been used successfully in the treatment of leprosy, nontuberculous mycobacteria, and visceral leishmaniasis.

Rigorous oral hygiene reduces but does not eliminate the discomfort of gingivitis, periodontal disease, and aphthous ulcers; chlorhexidine mouthwash and tooth brushing with a hydrogen peroxide–sodium bicarbonate paste also helps many patients. Oral antifungal agents (fluconazole, itraconazole, voriconazole, posaconazole) have reduced mucocutaneous candidiasis in patients with Job's syndrome. Androgens, glucocorticoids, lithium, and immunosuppressive therapy have been used to restore myelopoiesis in patients with neutropenia due to impaired production. Recombinant G-CSF is useful in the management of certain forms of neutropenia due to depressed neutrophil production, including those related to cancer chemotherapy. Patients with chronic neutropenia with evidence of a good bone marrow reserve need not receive prophylactic antibiotics. Patients with chronic or cyclic neutrophil counts <500/ μ L may benefit from prophylactic antibiotics and G-CSF during periods of neutropenia. Oral trimethoprim-sulfamethoxazole (160/800 mg) twice daily can prevent infection. Increased numbers of fungal infections are not seen in patients with CGD on this regimen. Oral quinolones such as levofloxacin and ciprofloxacin are alternatives.

In the setting of cytotoxic chemotherapy with severe, persistent lymphocyte dysfunction, trimethoprim-sulfamethoxazole prevents *Pneumocystis jiroveci* pneumonia. These patients, and patients with phagocytic cell dysfunction, should avoid heavy exposure to airborne soil, dust, or decaying matter (mulch, manure), which are often rich in *Nocardia* and the spores of *Aspergillus* and other fungi. Restriction of activities or social contact has no proven role in reducing risk of infection for phagocyte defects.

Although aggressive medical care for many patients with phagocytic disorders can allow them to go for years without a life-threatening infection, there may still be delayed effects of prolonged antimicrobials and other inflammatory complications. Cure of most congenital phagocyte defects is possible by bone marrow transplantation, and rates of success are improving ([Chap. 114](#)). The identification of specific gene defects in patients with LAD 1, CGD, and other immunodeficiencies has led to gene therapy trials in a number of genetic white cell disorders.

FURTHER READING

- BOELTZ S et al: To NET or not to NET: Current opinions and state of the science regarding the formation of neutrophil extracellular traps. *Cell Death Differ* 26:395, 2019.
- BOUSFIHA A et al: Human inborn errors of immunity: 2019 update of the IUIS phenotypical classification. *J Clin Immunol* 40:66, 2020.

DINAUER MC: Inflammatory consequences of inherited disorders affecting neutrophil function. *Blood* 133:2130, 2019.

KLION AD et al: Contributions of eosinophils to human health and disease. *Annu Rev Pathol* 15:179, 2020.

KUHNS DB: Diagnostic testing for chronic granulomatous disease. *Methods Mol Biol* 1982:543, 2019.

OCHOA S et al: Genetic susceptibility to fungal infection in children. *Curr Opin Pediatr* 32:780, 2020.

PEISELER M, KUBES P: More friend than foe: the emerging role of neutrophils in tissue repair. *J Clin Invest* 129:2629, 2019.

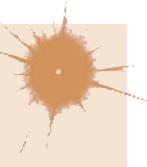
TANGYE SG et al: Human inborn errors of immunity: 2019 Update on the classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol* 40:24, 2020.

WU UI, HOLLAND SM: Host susceptibility to non-tuberculous mycobacterial infections. *Lancet Infect Dis* 15:968, 2015.

65

Bleeding and Thrombosis

Barbara A. Konkle



The human hemostatic system provides a natural balance between procoagulant and anticoagulant forces. The procoagulant forces include platelet adhesion and aggregation and fibrin clot formation; anticoagulant forces include the natural inhibitors of coagulation and fibrinolysis. Under normal circumstances, hemostasis is regulated to promote blood flow; however, it is also prepared to clot blood rapidly to arrest blood flow and prevent exsanguination. After bleeding is successfully halted, the system remodels the damaged vessel to restore normal blood flow. The major components of the hemostatic system, which function in concert, are (1) platelets and other formed elements of blood, such as monocytes and red cells; (2) plasma proteins (the coagulation and fibrinolytic factors and inhibitors); and (3) the vessel wall.

STEPS OF NORMAL HEMOSTASIS

■ PLATELET PLUG FORMATION

On vascular injury, platelets adhere to the site of injury, usually the denuded vascular intimal surface. Platelet adhesion is mediated primarily by von Willebrand factor (VWF), a large multimeric protein present in both plasma and the extracellular matrix of the subendothelial vessel wall, which serves as the primary "molecular glue," providing sufficient strength to withstand the high levels of shear stress that would tend to detach them with the flow of blood. Platelet adhesion is also facilitated by direct binding to subendothelial collagen through specific platelet membrane collagen receptors.

Platelet adhesion results in subsequent platelet activation and aggregation. This process is enhanced and amplified by humoral mediators in plasma (e.g., epinephrine, thrombin); mediators released from activated platelets (e.g., adenosine diphosphate, serotonin); and vessel wall extracellular matrix constituents that come in contact with adherent platelets (e.g., collagen, VWF). Activated platelets undergo the release reaction, during which they secrete contents that further promote aggregation and inhibit the naturally anticoagulant endothelial cell factors. During platelet aggregation (platelet-platelet interaction), additional platelets are recruited from the circulation to the site of vascular injury, leading to the formation of an occlusive platelet thrombus. The platelet plug is anchored and stabilized by the developing fibrin mesh.

The platelet glycoprotein (Gp) IIb/IIIa ($\alpha_{IIb} \beta_3$) complex is the most abundant receptor on the platelet surface. Platelet activation converts the normally inactive Gp IIb/IIIa receptor into an active receptor, enabling binding to fibrinogen and VWF. Because the surface of each platelet has about 50,000 Gp IIb/IIIa-binding sites, numerous activated

platelets recruited to the site of vascular injury can rapidly form an occlusive aggregate by means of a dense network of intercellular fibrinogen bridges.

FIBRIN CLOT FORMATION

Plasma coagulation proteins (*clotting factors*) normally circulate in plasma in their inactive forms. The sequence of coagulation protein reactions that culminate in the formation of fibrin was originally described as a *waterfall* or a *cascade*. Two pathways of blood coagulation have been described in the past: the so-called extrinsic, or tissue factor, pathway and the so-called intrinsic, or contact activation, pathway. We now know that coagulation is normally initiated through tissue factor (TF) exposure and activation through the classic *extrinsic pathway* but with critically important amplification through elements of the classic *intrinsic pathway*, as illustrated in Fig. 65-1. These reactions take place on phospholipid surfaces, usually the activated platelet surface. Coagulation testing in the laboratory can reflect other influences due to the artificial nature of the *in vitro* systems used (see below).

The immediate trigger for coagulation is vascular damage that exposes blood to TF that is constitutively expressed on the surfaces of subendothelial cellular components of the vessel wall, such as smooth muscle cells and fibroblasts. TF is also present in circulating microparticles, presumably shed from cells including monocytes and platelets. TF binds the serine protease factor VIIa; the complex activates factor X to factor Xa. Alternatively, the complex can indirectly activate factor X by initially converting factor IX to factor IXa, which then activates factor X. The participation of factor XI in hemostasis is not dependent on its activation by factor XIIa but rather on its positive feedback activation by thrombin. Thus, factor XIa functions in the propagation and amplification, rather than in the initiation, of the coagulation cascade. The role of factor XIIa in activation of factor XI is not fully elucidated, but studies suggest it may be a mechanism to promote thrombosis.

Factor Xa can be formed through the actions of either the TF/factor VIIa complex or factor IXa (with factor VIIIa as a cofactor) and converts prothrombin to thrombin, the pivotal protease of the coagulation system. The essential cofactor for this reaction is factor Va. Like the homologous factor VIIa, factor Va is produced by thrombin-induced limited proteolysis of factor V. Thrombin is a multifunctional enzyme that converts soluble plasma fibrinogen to an insoluble fibrin

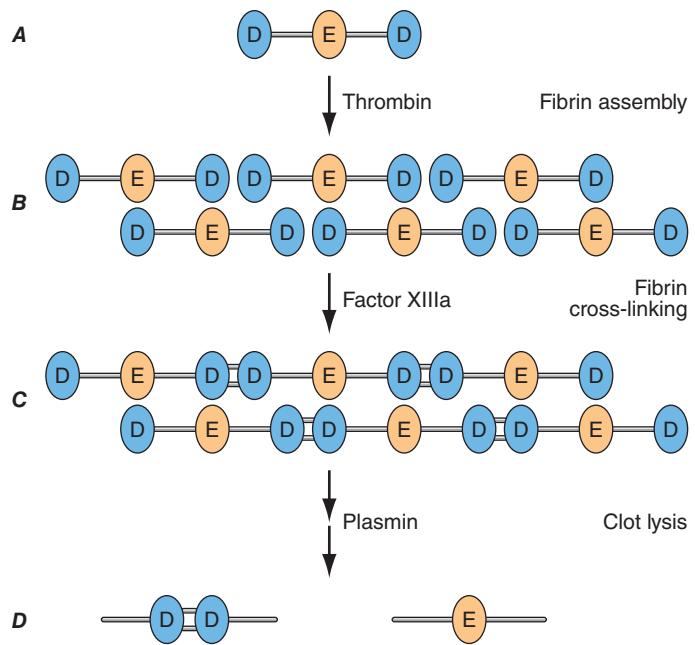


FIGURE 65-2 Fibrin formation and dissolution. (A) Fibrinogen is a trinodular structure consisting of two D domains and one E domain. Thrombin activation results in an ordered lateral assembly of protofibrils (B) with noncovalent associations. Factor XIIIa cross-links the D domains on adjacent molecules (C). Fibrin and fibrinogen (not shown) lysis by plasmin occurs at discrete sites and results in intermediary fibrin(ogen) degradation products (not shown). d-Dimers are the product of complete lysis of fibrin (D), maintaining the cross-linked D domains.

matrix. Fibrin polymerization involves an orderly process of intermolecular associations (Fig. 65-2). Thrombin also activates factor XIII (fibrin-stabilizing factor) to factor XIIIa, which covalently cross-links and thereby stabilizes the fibrin clot.

The assembly of the clotting factors on activated cell membrane surfaces greatly accelerates their reaction rates and also serves to localize blood clotting to sites of vascular injury. The critical cell membrane components, acidic phospholipids, are not normally exposed on resting cell membrane surfaces. However, when platelets, monocytes, and endothelial cells are activated by vascular injury or inflammatory stimuli, the procoagulant head groups of the membrane anionic phospholipids become translocated to the surfaces of these cells or released as part of microparticles, making them available to support and promote the plasma coagulation reactions.

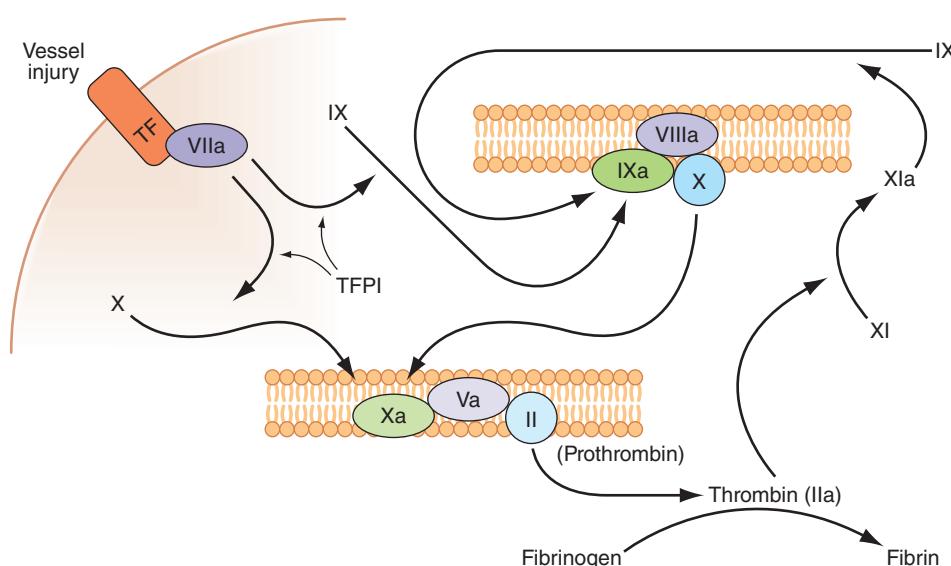


FIGURE 65-1 Coagulation is initiated by tissue factor (TF) exposure, which, with factor (F) VIIa, activates FIX and FX, which in turn, with FVIII and FV as cofactors, respectively, results in thrombin formation and subsequent conversion of fibrinogen to fibrin. Thrombin activates FXI, FVIII, and FV, amplifying the coagulation signal. Once the TF/VIIa/Fxa complex is formed, tissue factor pathway inhibitor (TFPI) inhibits the TF/VIIa pathway, making coagulation dependent on the amplification loop through FIX/FVIII. Coagulation requires calcium (not shown) and takes place on phospholipid surfaces, usually the activated platelet membrane.

ANTITHROMBOTIC MECHANISMS

Several physiologic antithrombotic mechanisms act in concert to prevent clotting under normal circumstances. These mechanisms operate to preserve blood fluidity and to limit blood clotting to specific focal sites of vascular injury. Endothelial cells have many antithrombotic effects. They produce prostacyclin, nitric oxide, and ectoADPase/CD39, which act to inhibit platelet binding, secretion, and aggregation. Endothelial cells produce anticoagulant factors including heparan proteoglycans, TF pathway inhibitor, and thrombomodulin. They also activate fibrinolytic mechanisms through the production of tissue plasminogen activator, urokinase, plasminogen activator inhibitors, and annexin-2.

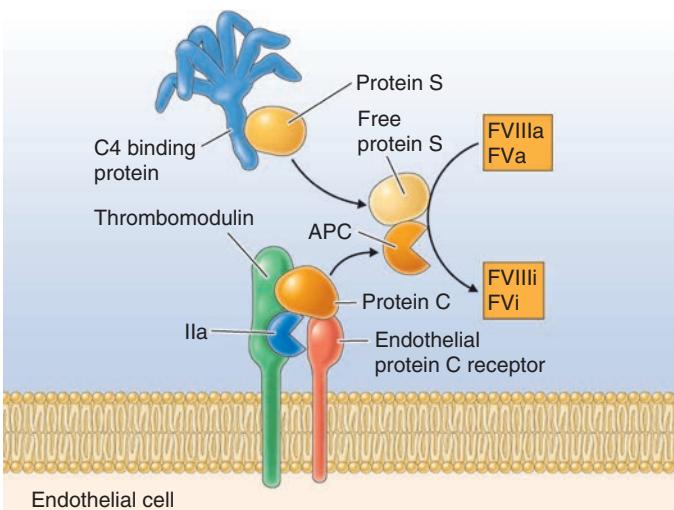


FIGURE 65-3 The activated protein C pathway in regulation of thrombosis. Thrombin generation results in protein C activation through interaction with thrombomodulin and protein C bound to the endothelial protein C receptor (EPCR). Activated protein C (APC) with free protein S converts activated factors (F) VIII and V to inactivate forms, thus in turn decreasing thrombin generation. C4BP, C4 binding protein; EC, endothelial cell; F, factor; IIa, thrombin; PC, protein C; PS, protein S; TM, thrombomodulin.

Antithrombin is the major plasma protease inhibitor of thrombin and the other clotting factors in coagulation. Antithrombin neutralizes thrombin and other activated coagulation factors by forming a complex between the active site of the enzyme and the reactive center of antithrombin. The rate of formation of these inactivating complexes increases by a factor of several thousand in the presence of heparin. Antithrombin inactivation of thrombin and other activated clotting factors occurs physiologically on vascular surfaces, where glycosaminoglycans, including heparan sulfates, are present to catalyze these reactions. Inherited quantitative or qualitative deficiencies of antithrombin lead to a lifelong predisposition to venous thromboembolism.

Protein C is a plasma glycoprotein that becomes an anticoagulant when it is activated by thrombin. The thrombin-induced activation of protein C occurs physiologically on thrombomodulin, a transmembrane proteoglycan-binding site for thrombin on endothelial cell surfaces. The binding of protein C to its receptor on endothelial cells places it in proximity to the thrombin-thrombomodulin complex, thereby enhancing its activation efficiency. (See Fig. 65-3.) Activated protein C acts as an anticoagulant by cleaving and inactivating activated factors V and VIII. This reaction is accelerated by a cofactor, protein S, which, like protein C, is a glycoprotein that undergoes vitamin K-dependent posttranslational modification. Quantitative or qualitative deficiencies of protein C or protein S, or resistance to the action of activated protein C by a specific variant at its target cleavage site in factor Va (factor V Leiden), lead to hypercoagulable states.

Tissue factor pathway inhibitor (TFPI) is a plasma protease inhibitor that regulates the TF-induced extrinsic pathway of coagulation. TFPI inhibits the TF/factor VIIa/factor Xa complex, essentially turning off the TF/factor VIIa initiation of coagulation, which then becomes dependent on the “amplification loop” via factor XI and factor VIII activation by thrombin. TFPI is bound to lipoprotein and can also be released by heparin from endothelial cells, where it is bound to glycosaminoglycans, and from platelets. The heparin-mediated release of TFPI may play a role in the anticoagulant effects of unfractionated and low-molecular-weight heparins.

THE FIBRINOLYTIC SYSTEM

Any thrombin that escapes the inhibitory effects of the physiologic anticoagulant systems is available to convert fibrinogen to fibrin. In response, the endogenous fibrinolytic system is then activated to

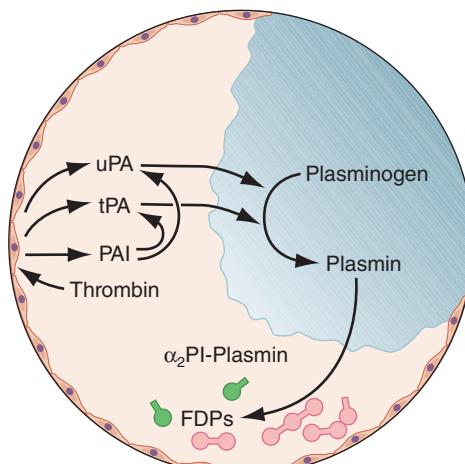


FIGURE 65-4 A schematic diagram of the fibrinolytic system. Tissue plasminogen activator (tPA) is released from endothelial cells, binds the fibrin clot, and activates plasminogen to plasmin. Release of plasminogen activator inhibitors (PAI-1 and PAI-2) inhibits tPA and urokinase (uPA). Excess fibrin is degraded by plasmin to distinct degradation products [FDPs (D-dimers)]. Any free plasmin is complexed with α_2 -antiplasmin (α_2 PI). PAI, plasminogen activator inhibitor; uPA, urokinase-type plasminogen activator.

dispose of intravascular fibrin and thereby maintain or reestablish the patency of the circulation. Just as thrombin is the key protease enzyme of the coagulation system, plasmin is the major protease enzyme of the fibrinolytic system, acting to digest fibrin to fibrin degradation products. The general scheme of fibrinolysis and its control is shown in Fig. 65-4.

The plasminogen activators, tissue type plasminogen activator (tPA) and the urokinase-type plasminogen activator (uPA), cleave the Arg560-Val561 bond of plasminogen to generate the active enzyme plasmin. The lysine-binding sites of plasmin (and plasminogen) permit it to bind to fibrin, so that physiologic fibrinolysis is “fibrin specific.” Both plasminogen (through its lysine-binding sites) and tPA possess specific affinity for fibrin and thereby bind selectively to clots. The assembly of a ternary complex, consisting of fibrin, plasminogen, and tPA, promotes the localized interaction between plasminogen and tPA and greatly accelerates the rate of plasminogen activation to plasmin. Moreover, partial degradation of fibrin by plasmin exposes new plasminogen and tPA-binding sites in carboxy-terminus lysine residues of fibrin fragments to enhance these reactions further. This creates a highly efficient mechanism to generate plasmin focally on the fibrin clot, which then becomes plasmin’s substrate for digestion to fibrin degradation products.

Plasmin cleaves fibrin at distinct sites of the fibrin molecule, leading to the generation of characteristic fibrin fragments during the process of fibrinolysis (Fig. 65-2). The sites of plasmin cleavage of fibrin are the same as those in fibrinogen. However, when plasmin acts on covalently cross-linked fibrin, D-dimers are released; hence, D-dimers can be measured in plasma as a relatively specific test of fibrin (rather than fibrinogen) degradation. D-Dimer assays can be used as sensitive markers of blood clot formation and have been validated for clinical use to exclude the diagnosis of deep venous thrombosis (DVT) and pulmonary embolism in selected populations. D-Dimer levels increase with age. A higher cut-off value to rule out venous thromboembolism (VTE) in the elderly has been proposed but is controversial.

Physiologic regulation of fibrinolysis occurs primarily at three levels: (1) plasminogen activator inhibitors (PAIs), specifically PAI-1 and PAI-2, inhibit the physiologic plasminogen activators; (2) the thrombin-activatable fibrinolysis inhibitor (TAFI) limits fibrinolysis; and (3) α_2 -antiplasmin inhibits plasmin. PAI-1 is the primary inhibitor of tPA and uPA in plasma. TAFI cleaves the N-terminal lysine residues of fibrin, which aid in localization of plasmin activity. α_2 -Antiplasmin is the main inhibitor of plasmin in human plasma, inactivating any nonfibrin clot-associated plasmin.

APPROACH TO THE PATIENT

Bleeding and Thrombosis

CLINICAL PRESENTATION

Disorders of hemostasis may be either inherited or acquired. A detailed personal and family history is key in determining the chronicity of symptoms and the likelihood of the disorder being inherited, as well as providing clues to underlying conditions that have contributed to the bleeding or thrombotic state. In addition, the history can give clues as to the etiology by determining (1) the bleeding (mucosal and/or joint) or thrombosis (arterial and/or venous) site and (2) whether an underlying bleeding or clotting tendency was enhanced by another medical condition or the introduction of medications or dietary supplements.

History of Bleeding A history of bleeding is the most important predictor of bleeding risk. In evaluating a patient for a bleeding disorder, a history of at-risk situations, including the response to past surgeries, should be assessed. Does the patient have a history of spontaneous or trauma/surgery-induced bleeding? Spontaneous hemarthroses are a hallmark of moderate and severe factor VIII and IX deficiency and, in rare circumstances, of other clotting factor deficiencies. Mucosal bleeding symptoms are more suggestive of underlying platelet disorders or von Willebrand disease (VWD), termed *disorders of primary hemostasis or platelet plug formation*. Disorders affecting primary hemostasis are shown in **Table 65-1**.

A bleeding score has been validated as a tool to predict patients more likely to have type 1 VWD (International Society on Thrombosis and Haemostasis Bleeding Assessment Tool [www.isth.org/resource/resmgr/ssc/isth-ssc_bleeding_assessment.pdf]), and a self-administered form has been validated. This is the most useful tool in excluding the diagnosis of a bleeding disorder, thus avoiding unnecessary testing, and is recommended by 2021 guidelines for screening for VWD in primary care. Bleeding symptoms that are more common in patients with bleeding disorders include prolonged bleeding with surgery, dental procedures and extractions, and/or trauma; heavy menstrual bleeding or postpartum hemorrhage; and large bruises (often described with lumps).

Easy bruising and heavy menstrual bleeding are common complaints in patients with and without bleeding disorders. Easy bruising can also be a sign of medical conditions in which there is no

identifiable coagulopathy; instead, the conditions are caused by an abnormality of blood vessels or their supporting connective tissues. In Ehlers-Danlos syndrome, there may be posttraumatic bleeding and a history of joint hyperextensibility. Cushing's syndrome, chronic steroid use, and aging result in changes in skin and subcutaneous tissue, and subcutaneous bleeding occurs in response to minor trauma. The latter has been termed *senile purpura*.

Epistaxis is a common symptom, particularly in children and in dry climates, and may not reflect an underlying bleeding disorder. However, it is the most common symptom in hereditary hemorrhagic telangiectasia and in boys with VWD. Clues that epistaxis is a symptom of an underlying bleeding disorder include lack of seasonal variation and bleeding that requires medical evaluation or treatment, including cauterization. Bleeding with eruption of primary teeth is seen in children with more severe bleeding disorders, such as moderate and severe hemophilia. It is uncommon in children with mild bleeding disorders. Patients with disorders of primary hemostasis (platelet adhesion) may have increased bleeding after dental cleanings and other procedures that involve gum manipulation.

Heavy menstrual bleeding is defined quantitatively as a loss of >80 mL of blood per cycle, based on the quantity of blood loss required to produce iron-deficiency anemia. A complaint of heavy menses is subjective and has a poor correlation with excessive blood loss. Predictors of heavy menstrual bleeding include bleeding resulting in iron-deficiency anemia or a need for blood transfusion, passage of clots >1 inch in diameter, and changing a pad or tampon more than hourly. Heavy menstrual bleeding is a common symptom in women with underlying bleeding disorders and is reported in the majority of women with VWD, factor XI deficiency, platelet function disorders, and hemophilia, including genetic carriers with borderline-normal factor levels. Women with underlying bleeding disorders are more likely to have other bleeding symptoms, including bleeding after dental extractions and postoperative and postpartum bleeding, and are much more likely to have heavy menstrual bleeding beginning at menarche than women with heavy menstrual bleeding due to other causes. Heavy menstrual bleeding may result in iron deficiency and is documented to have significant adverse effects on quality of life.

Postpartum hemorrhage is a common symptom in women with underlying bleeding disorders. In women with type 1 VWD or hemophilia A in whom levels of VWF and factor VIII usually normalize during pregnancy, postpartum hemorrhage may be delayed. Women with a history of postpartum hemorrhage may have a higher risk of recurrence with subsequent pregnancies. Women with underlying bleeding disorders are at risk for other reproductive tract bleeding, including rupture of ovarian cysts with intraabdominal hemorrhage.

Tonsillectomy is a major hemostatic challenge, because intact hemostatic mechanisms are essential to prevent excessive bleeding from the tonsillar bed. Bleeding may occur early after surgery or after approximately 7 days postoperatively, with loss of the eschar at the operative site. Similar delayed bleeding is seen after colonic polyp resection. Gastrointestinal (GI) bleeding and hematuria are usually due to underlying pathology, and procedures to identify and treat the bleeding site should be undertaken, even in patients with known bleeding disorders. VWD, particularly types 2 and 3, is associated with angiodyplasia of the bowel and GI bleeding.

Hemarthroses and spontaneous muscle hematomas are characteristic of moderate or severe congenital factor VIII or IX deficiency. They can also be seen in moderate and severe deficiencies of fibrinogen, prothrombin, and factors V, VII, and X. Spontaneous hemarthroses occur rarely in other bleeding disorders except for severe VWD, with associated factor VIII levels <5%. Muscle and soft tissue bleeds are also common in acquired factor VIII deficiency. Bleeding into a joint results in severe pain and swelling, as well as loss of function, but is rarely associated with discoloration from bruising around the joint. Life-threatening sites of bleeding

TABLE 65-1 Primary Hemostatic (Platelet Plug) Disorders

Defects of Platelet Adhesion

von Willebrand disease

Bernard-Soulier syndrome (absence or dysfunction of platelet Gp Ib-IX-V)

Defects of Platelet Aggregation

Glanzmann's thrombasthenia (absence or dysfunction of platelet glycoprotein [Gp] IIb/IIIa)

Afibrinogenemia

Defects of Platelet Secretion

Decreased cyclooxygenase activity

Drug-induced (aspirin, nonsteroidal anti-inflammatory agents, thienopyridines)

Inherited

Acquired

Granule storage pool defects

Inherited

Acquired

Nonspecific inherited secretory defects

Nonspecific drug effects

Uremia

Platelet coating (e.g., paraprotein, penicillin)

Defect of Platelet Coagulant Activity

Scott's syndrome

include bleeding into the oropharynx, where bleeding can obstruct the airway, into the central nervous system, and into the retroperitoneum. Central nervous system bleeding is the major cause of bleeding-related deaths in patients with severe congenital factor deficiencies.

Prohemorrhagic Effects of Medications and Dietary Supplements

Aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) that inhibit cyclooxygenase 1 impair primary hemostasis and may exacerbate bleeding from another cause or even unmask a previously occult mild bleeding disorder such as VWD. All NSAIDs, however, can precipitate GI bleeding, which may be more severe in patients with underlying bleeding disorders. The aspirin effect on platelet function lasts for the life of the platelet; however, in individuals with typical platelet turnover, the functional defect reverts to near-normal within 2–3 days after the last dose. The effect of other NSAIDs is shorter, as the inhibitor effect is reversed when the drug is removed. Inhibitors of the ADP P2Y₁₂ receptor (clopidogrel, prasugrel, and ticagrelor) inhibit ADP-mediated platelet aggregation and, like NSAIDs, can precipitate or exacerbate bleeding symptoms. The risk of bleeding with these drugs is higher than with NSAIDs.

Many herbal supplements can impair hemostatic function (**Table 65-2**). Some are more convincingly associated with a bleeding risk than others. Fish oil or concentrated omega-3 fatty acid supplements impair platelet function. They alter platelet biochemistry to produce more PGI₃, a more potent platelet inhibitor than prostacyclin (PGI₂), and more thromboxane A₃, a less potent platelet activator than thromboxane A₂. In fact, diets naturally rich in omega-3 fatty acids can result in a prolonged bleeding time and abnormal platelet aggregation studies, but the actual associated bleeding risk is unclear. Vitamin E appears to inhibit protein kinase C-mediated platelet aggregation and nitric oxide production. In patients with unexplained bruising or bleeding, it is prudent to review any new medications or supplements and discontinue those that may be associated with bleeding.

Underlying Systemic Diseases That Cause or Exacerbate a Bleeding Tendency

Acquired bleeding disorders are commonly secondary to, or associated with, systemic disease. The clinical evaluation of a patient with a bleeding tendency must therefore include a thorough assessment for evidence of underlying disease. Bruising or mucosal bleeding may be the presenting complaint in

liver disease, severe renal impairment, hypothyroidism, paraproteinemias or amyloidosis, and conditions causing bone marrow failure. All coagulation factors are synthesized in the liver, and hepatic failure results in combined factor deficiencies. This is often compounded by thrombocytopenia and portal hypertension. Coagulation factors II, VII, IX, and X and proteins C, S, and Z are dependent on vitamin K for posttranslational modification. Although vitamin K is required in both procoagulant and anticoagulant processes, the phenotype of vitamin K deficiency or the warfarin effect on coagulation is bleeding.

The normal blood platelet count is 150,000–450,000/ μ L. Thrombocytopenia results from decreased production, increased destruction, and/or sequestration. Although the bleeding risk varies somewhat by the reason for the thrombocytopenia, bleeding rarely occurs in isolated thrombocytopenia at counts >50,000/ μ L and usually not until <10,000–20,000/ μ L. Coexisting coagulopathies, as is seen in liver failure or disseminated coagulation; infection; platelet-inhibitory drugs; and underlying medical conditions can all increase the risk of bleeding in the thrombocytopenic patient. Most procedures can be performed in patients with a platelet count of 50,000/ μ L or greater.

HISTORY OF THROMBOSIS

The risk of thrombosis, like that of bleeding, is influenced by both genetic and environmental factors. The major risk factor for arterial thrombosis is atherosclerosis, whereas for venous thrombosis, the risk factors are immobility, surgery, underlying medical conditions such as malignancy, medications such as hormonal therapy, obesity, and genetic predispositions. Factors that increase risks for venous and for both venous and arterial thromboses are shown in **Table 65-3**.

The most important point in a history related to venous thrombosis is determining whether the thrombotic event was idiopathic (meaning there was no clear precipitating factor) or was a precipitated event. In patients without underlying malignancy, having an idiopathic event is the strongest predictor of recurrence of VTE. In patients who have a vague history of thrombosis, a history of being treated with warfarin or other anticoagulants suggests a past DVT. Age is an important risk factor for venous thrombosis—the risk of DVT increases per decade, with an approximate incidence of 1/100,000 per year in early childhood to 1/200 per year among octogenarians. Family history is helpful in determining if there is a

TABLE 65-2 Herbal Supplements Associated with Increased Bleeding

Herbs with Potential Antiplatelet Activity

- Ginkgo (*Ginkgo biloba L.*)
- Garlic (*Allium sativum*)
- Bilberry (*Vaccinium myrtillus*)
- Ginger (*Gingiber officinale*)
- Dong quai (*Angelica sinensis*)
- Feverfew (*Tanacetum parthenium*)
- Asian ginseng (*Panax ginseng*)
- American ginseng (*Panax quinquefolius*)
- Siberian ginseng/eleuthero (*Eleutherococcus senticosus*)
- Turmeric (*Circuma longa*)
- Meadowsweet (*Filipendula ulmaria*)
- Willow (*Salix spp.*)

Coumarin-Containing Herbs

- Motherwort (*Leonurus cardiaca*)
- Chamomile (*Matricaria recutita, Chamaemelum mobile*)
- Horse chestnut (*Aesculus hippocastanum*)
- Red clover (*Trifolium pratense*)
- Fenugreek (*Trigonella foenum-graecum*)

TABLE 65-3 Some Risk Factors for Thrombosis

VENOUS	VENOUS AND ARTERIAL
Inherited	Inherited
Factor V Leiden	Homocystinuria
Prothrombin G20210A	Dysfibrinogenemia
Antithrombin deficiency	
Protein C deficiency	
Protein S deficiency	
Acquired	Acquired
Age	Malignancy
Previous thrombosis	Antiphospholipid antibody syndrome
Immobilization	Hormonal therapy
Major surgery	Polycythemia vera
Pregnancy and puerperium	Essential thrombocythemia
Hospitalization	Paroxysmal nocturnal hemoglobinuria
Obesity	Thrombotic thrombocytopenic purpura
Infection	Heparin-induced thrombocytopenia
Smoking	Disseminated intravascular coagulation
	Unknown^a
	Elevated factor II, VIII, IX, XI
	Elevated TAFI levels
	Low levels of TFPI

^aUnknown whether risk is inherited or acquired.

Abbreviations: APC, activated protein C; TAFI, thrombin-activatable fibrinolysis inhibitor; TFPI, tissue factor pathway inhibitor.

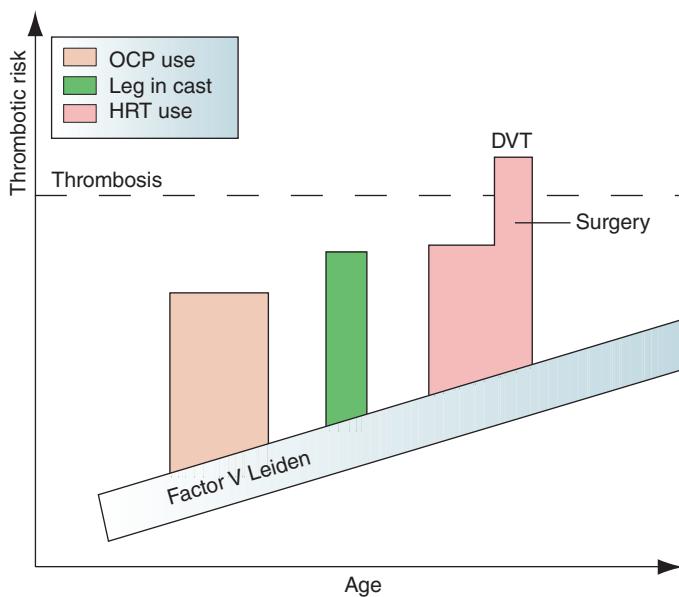


FIGURE 65-5 Thrombotic risk over time. Shown schematically is an individual's thrombotic risk over time. An underlying factor V Leiden variant provides a "theoretically" constant increased risk. The thrombotic risk increases with age and, intermittently, with oral contraceptive (OCP) or hormone replacement therapy (HRT) use; other events, like major surgery or illness, will increase the risk further. At some point, the cumulative risk may increase to the threshold for thrombosis and result in deep venous thrombosis (DVT). Note: The magnitude and duration of risk portrayed in the figure are meant for example only and may not precisely reflect the relative risk determined by clinical study. (Sources: From BA Konkle, A Schafer, in DP Zipes et al [eds]: *Braunwald's Heart Disease*, 7th ed. Philadelphia, Saunders, 2005; from FR Rosendaal: Venous thrombosis: A multicausal disease. *Lancet* 353:1167, 1999.)

genetic predisposition and how strong that predisposition appears to be. A genetic thrombophilia that confers a relatively small increased risk, such as being a heterozygote for the prothrombin G20210A or factor V Leiden mutation, is a minor determinant of risk in an elderly individual undergoing a high-risk surgical procedure. As illustrated in Fig. 65-5, a thrombotic event usually has more than one contributing factor. Predisposing factors must be carefully assessed to determine the risk of recurrent thrombosis and, with consideration of the patient's bleeding risk, determine the length of anticoagulation. Testing for inherited thrombophilias in adults should be limited to instances where results would change clinical care. Such instances are rare.

LABORATORY EVALUATION

Careful history taking and clinical examination are essential components in the assessment of bleeding and thrombotic risk. The use of laboratory tests of coagulation complements, but cannot substitute for, clinical assessment. No test exists that provides a global assessment of hemostasis. The bleeding time has been used to assess bleeding risk; however, it does not predict bleeding risk with surgery, and it is not recommended for this indication. The PFA-100, an instrument that measures platelet-dependent coagulation under flow conditions, is more sensitive and specific for VWD than the bleeding time; however, it is not sensitive enough to rule out mild bleeding disorders. PFA-100 closure times are prolonged in patients with some, but not all, inherited platelet disorders. Also, its utility in predicting bleeding risk has not been determined. Thromboelastography can be useful in guiding intraoperative transfusion and is being explored in other settings, but is not broadly applicable for the diagnosis of disorders of hemostasis and thrombosis.

For routine preoperative and preprocedure testing, an abnormal prothrombin time (PT) may detect liver disease or vitamin K deficiency that had not been previously appreciated. Studies have not confirmed the usefulness of an activated partial thromboplastin time (aPTT) in preoperative evaluations in patients with a negative

bleeding history. The primary use of coagulation testing should be to confirm the presence and type of bleeding disorder in a patient with a suspicious clinical history.

Because of the nature of coagulation assays, proper sample acquisition and handling is critical to obtaining valid results. In patients with abnormal coagulation assays who have no bleeding history, repeat studies with attention to these factors frequently results in normal values. Most coagulation assays are performed in sodium citrate anticoagulated plasma that is recalcified for the assay. Because the anticoagulant is in liquid solution and needs to be added to blood in proportion to the plasma volume, incorrectly filled or inadequately mixed blood collection tubes will give erroneous results. These vacutainer tubes should be filled to >90% of the recommended fill, which is usually denoted by a line on the tube. An elevated hematocrit (>55%) can result in a false value due to a decreased plasma-to-anticoagulant ratio.

Screening Assays The most commonly used screening tests are the PT, aPTT, and platelet count. The PT assesses the factors I (fibrinogen), II (prothrombin), V, VII, and X (Fig. 65-6). The PT measures the time for clot formation of the citrated plasma after recalcification and addition of thromboplastin, a mixture of TF and phospholipids. The sensitivity of the assay varies by the source of thromboplastin. The relationship between defects in secondary hemostasis (fibrin formation) and coagulation test abnormalities is shown in Table 65-4. To adjust for this variability, the overall sensitivity of different thromboplastins to reduction of the vitamin K-dependent clotting factors II, VII, IX, and X in anticoagulation patients is expressed as the International Sensitivity Index (ISI). The international normalized ratio (INR) is determined based on the formula: $INR = (\text{PT}^{\text{patient}}/\text{PT}^{\text{normal mean}})^{\text{ISI}}$.

The INR was developed to assess stable anticoagulation due to reduction of vitamin K-dependent coagulation factors; it is commonly used in the evaluation of patients with liver disease. Although it does allow comparison between laboratories, reagent sensitivity as used to determine the ISI is not the same in liver disease as with warfarin anticoagulation. In addition, progressive liver failure is associated with variable changes in coagulation factors; the degree of prolongation of either the PT or the INR only roughly

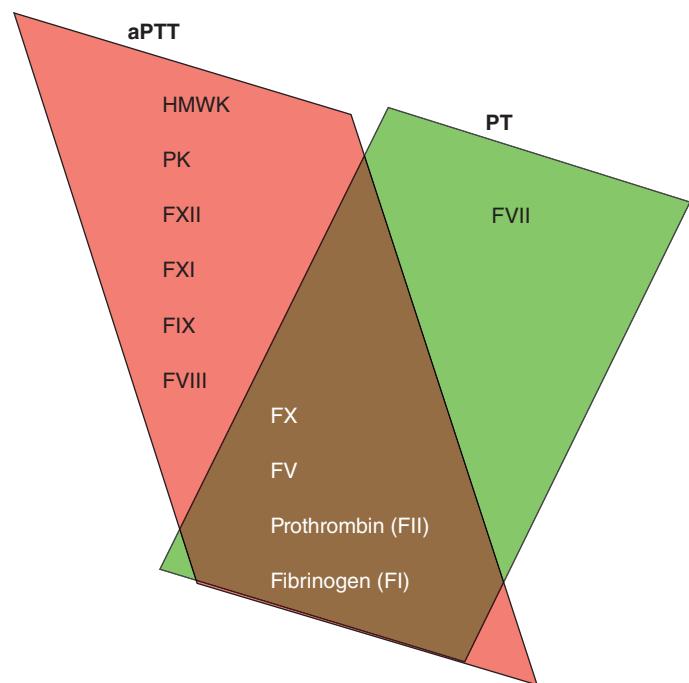


FIGURE 65-6 Coagulation factor activity tested in the activated partial thromboplastin time (aPTT) in red and prothrombin time (PT) in green, or both. F, factor; HMWK, high-molecular-weight kininogen; PK, prekallikrein.

TABLE 65-4 Hemostatic Disorders and Coagulation Test Abnormalities	
Prolonged Activated Partial Thromboplastin Time (aPTT)	
No clinical bleeding—↓ factor XII, high-molecular-weight kininogen, prekallikrein	
Variable, but usually mild, bleeding—↓ factor XI, mild ↓ factor VIII and factor IX	
Frequent, severe bleeding—severe deficiencies of factors VIII and IX	
Heparin and direct thrombin inhibitors	
Prolonged Prothrombin Time (PT)	
Factor VII deficiency	
Vitamin K deficiency—early	
Warfarin anticoagulation	
Direct Xa inhibitors (rivaroxaban, edoxaban, apixaban—note PT may be normal)	
Prolonged aPTT and PT	
Factor II, V, X, or fibrinogen deficiency	
Vitamin K deficiency—late	
Direct thrombin inhibitors	
Prolonged Thrombin Time	
Heparin or heparin-like inhibitors	
Direct thrombin inhibitors (e.g., dabigatran, argatroban, bivalirudin)	
Mild or no bleeding—dysfibrinogenemia	
Frequent, severe bleeding—afibrinogenemia	
Prolonged PT and/or aPTT Not Corrected with Mixing with Normal Plasma	
Bleeding—specific factor inhibitor	
No symptoms, or clotting and/or pregnancy loss—lupus anticoagulant	
Disseminated intravascular coagulation	
Heparin or direct thrombin inhibitor	
Abnormal Clot Solubility	
Factor XIII deficiency	
Inhibitors or defective cross-linking	
Rapid Clot Lysis	
Deficiency of α_2 -antiplasmin or plasminogen activator inhibitor 1	
Treatment with fibrinolytic therapy	

predicts the bleeding risk. Thrombin generation has been shown to be normal in many patients with mild to moderate liver dysfunction. Because the PT only measures one aspect of hemostasis affected by liver dysfunction, we likely overestimate the bleeding risk of a mildly elevated INR in this setting. PT reagents have variable sensitivity to the direct Xa inhibitors, and the PT is usually normal in patients on apixaban.

The aPTT assesses the intrinsic and common coagulation pathways; factors XI, IX, VIII, X, V, and II; fibrinogen; prekallikrein; high-molecular-weight kininogen; and factor XII (Fig. 65-6). The aPTT reagent contains phospholipids derived from either animal or vegetable sources that function as a platelet substitute in the coagulation pathways and includes an activator of the intrinsic coagulation system, such as nonparticulate ellagic acid or the particulate activators kaolin, celite, or micronized silica.

The phospholipid composition of aPTT reagents varies, which influences the sensitivity of individual reagents to clotting factor deficiencies and to inhibitors such as heparin and lupus anticoagulants. Thus, aPTT results will vary from one laboratory to another, and the normal range in the laboratory where the testing occurs should be used in the interpretation. Local laboratories can relate their aPTT values to the therapeutic heparin anticoagulation by correlating aPTT values with direct measurements of heparin activity (anti-Xa or protamine titration assays) in samples from heparinized patients, although correlation between these assays is

often poor. The aPTT reagent will vary in sensitivity to individual factor deficiencies and usually becomes prolonged with individual factor deficiencies of 30–50%.

Mixing Studies Mixing studies are used to evaluate a prolonged aPTT or, less commonly PT, to distinguish between a factor deficiency and an inhibitor. In this assay, normal plasma and patient plasma are mixed in a 1:1 ratio, and the aPTT or PT is determined immediately and after incubation at 37°C for varying times, typically 30, 60, and/or 120 min. With isolated factor deficiencies, the aPTT will correct with mixing and stay corrected with incubation. With aPTT prolongation due to a lupus anticoagulant, the mixing and incubation will show no correction. In acquired neutralizing factor antibodies, notably an acquired factor VIII inhibitor, the initial assay may or may not correct immediately after mixing but will prolong or remain prolonged with incubation at 37°C. Failure to correct with mixing can also be due to the presence of other inhibitors or interfering substances such as heparin, fibrin split products, and paraproteins.

Specific Factor Assays Decisions to proceed with specific clotting factor assays will be influenced by the clinical situation and the results of coagulation screening tests. Precise diagnosis and effective management of inherited and acquired coagulation deficiencies necessitate quantitation of the relevant factors. When bleeding is severe, specific assays are urgently required to guide appropriate therapy. Individual factor assays are usually performed as modifications of the mixing study, where the patient's plasma is mixed with plasma deficient in the factor being studied. This will correct all factor deficiencies to >50%, thus making prolongation of clot formation due to a factor deficiency dependent on the factor missing from the added plasma.

Testing for Antiphospholipid Antibodies Antibodies to phospholipids (cardiolipin) or phospholipid-binding proteins (β_2 -microglobulin and others) are detected by enzyme-linked immunosorbent assay (ELISA). When these antibodies interfere with phospholipid-dependent coagulation tests, they are termed *lupus anticoagulants*. The aPTT has variability sensitivity to lupus anticoagulants, depending in part on the aPTT reagents used. An assay using a sensitive reagent has been termed an *LA-PTT*. The dilute Russell viper venom test (dRVVT) is a modification of a standard test with the phospholipid reagent decreased, thus increasing the sensitivity to antibodies that interfere with the phospholipid component. These tests, however, are not specific for lupus anticoagulants, because factor deficiencies or other inhibitors will also result in prolongation. Documentation of a lupus anticoagulant requires not only prolongation of a phospholipid-dependent coagulation test but also lack of correction when mixed with normal plasma and correction with the addition of activated platelet membranes or certain phospholipids (e.g., hexagonal phase).

Other Coagulation Tests The thrombin time and the reptilase time measure fibrinogen conversion to fibrin and are prolonged when the fibrinogen level is low (usually <80–100 mg/dL) or qualitatively abnormal, as seen in inherited or acquired dysfibrinogenemias, or when fibrin/fibrinogen degradation products interfere. The thrombin time, but not the reptilase time, is prolonged in the presence of heparin. The thrombin time is markedly prolonged in the presence of the direct thrombin inhibitor, dabigatran; a dilute thrombin time can be used to assess drug activity. Measurement of anti-factor Xa plasma inhibitory activity is a test frequently used to assess low-molecular-weight heparin (LMWH) levels, as a direct measurement of unfractionated heparin (UFH) activity, or to assess activity of the direct Xa inhibitors rivaroxaban, apixaban, and edoxaban. Drug in the patient sample inhibits the enzymatic conversion of an Xa-specific chromogenic substrate to colored product by factor Xa. Standard curves are created using multiple concentrations of the specific drug and are used to calculate the concentration of anti-Xa activity in the patient plasma.

Laboratory Testing for Thrombophilia Laboratory assays to detect thrombophilic states include molecular diagnostics and immuno-logic and functional assays. These assays vary in their sensitivity and specificity for the condition being tested. Furthermore, acute thrombosis, acute illnesses, inflammatory conditions, pregnancy, and medications affect levels of many coagulation factors and their inhibitors. Antithrombin is decreased by heparin and in the setting of acute thrombosis. Protein C and S levels may be increased in the setting of acute thrombosis and are decreased by warfarin. Antiphospholipid antibodies are frequently transiently positive in acute illness. Testing for genetic thrombophilias should, in general, only be performed when there is a strong family history of thrombosis and results would affect clinical decision-making.

Because thrombophilia evaluations are usually performed to assess the need to extend anticoagulation, testing, if indicated, should be performed in a steady state, remote from the acute event. Functional assays, but not genetic assays, will be affected by anti-coagulants including warfarin (for vitamin K-dependent proteins) and thrombin and Xa inhibitors and cannot be interpreted in patients on those drugs. In most instances, when discontinuation of anticoagulation is being considered, drugs can be stopped after the initial 3–6 months of treatment, and testing can be performed at least 3 weeks later.

Measures of Platelet Function The bleeding time was used in the past to assess bleeding risk; however, it has not been found to predict bleeding risk with surgery, and it is not recommended for use for this indication. The PFA-100 and similar instruments that measure platelet-dependent coagulation under flow conditions are generally more sensitive and specific for platelet disorders and VWD than the bleeding time; however, data are insufficient to support their use to predict bleeding risk or monitor response to therapy, and they will be normal in some patients with platelet disorders or mild VWD. When they are used in the evaluation of a patient with bleeding symptoms, abnormal results require specific testing, such as VWF assays and/or platelet aggregation studies. Because all of these “screening” assays may miss patients with mild bleeding disorders, further studies are needed to define their role in hemostasis testing.

For classic platelet aggregometry, various agonists are added to the patient’s platelet-rich plasma or whole blood, and platelet aggregation is measured. Tests of platelet secretion in response to agonists can also be measured. These remain the gold standard for diagnosis of platelet function disorders. However, they are affected by many factors, including numerous medications, and the association between minor defects in these assays and bleeding risk is not clearly established.

FURTHER READING

- CHAPIN JC, HAJJAR KA: Fibrinolysis and the control of blood coagulation. *Blood Rev* 29:17, 2015.
- CONNORS JM: Thrombophilia testing and venous thrombosis. *N Engl J Med* 377:12, 2017.
- CONNORS JM: Testing and monitoring direct oral anticoagulants. *Blood* 132:2009, 2018.
- DARZI AJ et al: Prognostic factors for VTE and bleeding in hospitalized medical patients: A systematic review and meta-analysis. *Blood* 135:1788, 2020.
- DEVREESE KMJ et al: Guidance from the Scientific and Standardization Committee for lupus anticoagulant/antiphospholipid antibodies of the International Society on Thrombosis and Haemostasis Update of the guidelines for lupus anticoagulant detection and interpretation. *J Thromb Haemost* 18:2828, 2020.
- ELBAZ C, SHOLZBERG M: An illustrated review of bleeding assessment tools and common coagulation tests. *Res Pract Thromb Haemost* 4:761, 2020.
- JAMES PD et al: ASH ISTH NHF WFH 2021 guidelines on the diagnosis of von Willebrand disease. *Blood Adv* 5:280, 2021.

KAUFMAN RM et al: Platelet transfusion: A clinical practice guideline from the AABB. *Ann Intern Med* 162:205, 2020.

MACKIE I et al: Guidelines on the laboratory aspect of assays used in haemostasis and thrombosis. *Int J Lab Hem* 35:1, 2013.

MORAN J, BAUER KA: Managing thromboembolic risk in patients with hereditary and acquired thrombophilias. *Blood* 135:344, 2020.

WAGENMAN BL et al: The laboratory approach to inherited and acquired coagulation factor deficiencies. *Clin Lab Med* 29:229, 2009.

YAU JW et al: Endothelial cell control of thrombosis. *BMC Cardiovasc Disord* 15:130, 2015.

66

Enlargement of Lymph Nodes and Spleen

Dan L. Longo



This chapter is intended to serve as a guide to the evaluation of patients who present with enlargement of the lymph nodes (*lymphadenopathy*) or the spleen (*splenomegaly*). Lymphadenopathy is a rather common clinical finding in primary care settings, whereas palpable splenomegaly is less so.

LYMPHADENOPATHY

Lymphadenopathy may be an incidental finding in patients being examined for various reasons, or it may be a presenting sign or symptom of the patient’s illness. The physician must eventually decide whether the lymphadenopathy is a normal finding or one that requires further study, up to and including biopsy. Soft, flat, submandibular nodes (<1 cm) are often palpable in healthy children and young adults; healthy adults may have palpable inguinal nodes of up to 2 cm, which are considered normal. Further evaluation of these normal nodes is not warranted. In contrast, if the physician believes the node(s) to be abnormal, then pursuit of a more precise diagnosis is needed.

APPROACH TO THE PATIENT

Lymphadenopathy

Lymphadenopathy may be a primary or secondary manifestation of numerous disorders, as shown in **Table 66-1**. Many of these disorders are infrequent causes of lymphadenopathy. In primary care practice, more than two-thirds of patients with lymphadenopathy have nonspecific causes or upper respiratory illnesses (viral or bacterial) and <1% have a malignancy. In one study, 84% of patients referred for evaluation of lymphadenopathy had a “benign” diagnosis. The remaining 16% had a malignancy (lymphoma or metastatic adenocarcinoma). Of the patients with benign lymphadenopathy, 63% had a nonspecific or reactive etiology (no causative agent found), and the remainder had a specific cause demonstrated, most commonly infectious mononucleosis, toxoplasmosis, or tuberculosis. Thus, the vast majority of patients with lymphadenopathy will have a nonspecific etiology requiring few diagnostic tests.

CLINICAL ASSESSMENT

The physician will be aided in the pursuit of an explanation for the lymphadenopathy by a careful medical history, physical examination, selected laboratory tests, and perhaps an excisional lymph node biopsy.

The *medical history* should reveal the setting in which lymphadenopathy is occurring. Symptoms such as sore throat, cough, fever, night sweats, fatigue, weight loss, or pain in the nodes should be

TABLE 66-1 Diseases Associated with Lymphadenopathy

1. Infectious diseases
 - a. Viral—*infectious mononucleosis syndromes (EBV, CMV), infectious hepatitis, herpes simplex, herpesvirus-6, varicella-zoster virus, rubella, measles, adenovirus, HIV, epidemic keratoconjunctivitis, vaccinia, herpesvirus-8*
 - b. Bacterial—*streptococci, staphylococci, cat-scratch disease, brucellosis, tularemia, plague, chancroid, melioidosis, glanders, tuberculosis, atypical mycobacterial infection, primary and secondary syphilis, diphtheria, leprosy, bartonella*
 - c. Fungal—*histoplasmosis, coccidioidomycosis, paracoccidioidomycosis*
 - d. Chlamydial—*lymphogranuloma venereum, trachoma*
 - e. Parasitic—*toxoplasmosis, leishmaniasis, trypanosomiasis, filariasis*
 - f. Rickettsial—*scrub typhus, rickettsialpox, Q fever*
2. Immunologic diseases
 - a. Rheumatoid arthritis
 - b. Juvenile rheumatoid arthritis
 - c. Mixed connective tissue disease
 - d. Systemic lupus erythematosus
 - e. Dermatomyositis
 - f. Sjögren's syndrome
 - g. Serum sickness
 - h. Drug hypersensitivity—*diphenylhydantoin, hydralazine, allopurinol, primidone, gold, carbamazepine, etc.*
 - i. Angioimmunoblastic lymphadenopathy
 - j. Primary biliary cirrhosis
 - k. Graft-vs-host disease
 - l. Silicone-associated
 - m. Autoimmune lymphoproliferative syndrome
 - n. IgG4-related disease
 - o. Immune reconstitution inflammatory syndrome (IRIS)
3. Malignant diseases
 - a. Hematologic—*Hodgkin's disease, non-Hodgkin's lymphomas, acute or chronic lymphocytic leukemia, hairy cell leukemia, malignant histiocytosis, amyloidosis*
 - b. Metastatic—from numerous primary sites
4. Lipid storage diseases—*Gaucher's, Niemann-Pick, Fabry, Tangier*
5. Endocrine diseases—*hyperthyroidism*
6. Other disorders
 - a. Castleman's disease (giant lymph node hyperplasia)
 - b. Sarcoidosis
 - c. Dermatopathic lymphadenitis
 - d. Lymphomatoid granulomatosis
 - e. Histiocytic necrotizing lymphadenitis (Kikuchi's disease)
 - f. Sinus histiocytosis with massive lymphadenopathy (Rosai-Dorfman disease)
 - g. Mucocutaneous lymph node syndrome (Kawasaki's disease)
 - h. Histiocytosis X
 - i. Familial Mediterranean fever
 - j. Severe hypertriglyceridemia
 - k. Vascular transformation of sinuses
 - l. Inflammatory pseudotumor of lymph node
 - m. Congestive heart failure

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus.

sought. The patient's age, sex, occupation, exposure to pets, sexual behavior, and use of drugs such as diphenylhydantoin are other important historic points. For example, children and young adults usually have benign (i.e., nonmalignant) disorders that account for the observed lymphadenopathy such as viral or bacterial upper respiratory infections; infectious mononucleosis; toxoplasmosis; and, in some countries, tuberculosis. In contrast, after age 50, the incidence of malignant disorders increases and that of benign disorders decreases.

The physical examination can provide useful clues such as the extent of lymphadenopathy (localized or generalized), size of nodes, texture, presence or absence of nodal tenderness, signs of inflammation over the node, skin lesions, and splenomegaly. A thorough ear, nose, and throat (ENT) examination is indicated in adult patients with cervical adenopathy and a history of tobacco use. Localized or regional adenopathy implies involvement of a single anatomic area. Generalized adenopathy has been defined as involvement of three or more noncontiguous lymph node areas. Many of the causes of lymphadenopathy (Table 66-1) can produce localized or generalized adenopathy, so this distinction is of limited utility in the differential diagnosis. Nevertheless, generalized lymphadenopathy is frequently associated with nonmalignant disorders such as infectious mononucleosis (Epstein-Barr virus [EBV] or cytomegalovirus [CMV]), toxoplasmosis, AIDS, other viral infections, systemic lupus erythematosus (SLE), and mixed connective tissue disease. Acute and chronic lymphocytic leukemias and malignant lymphomas also produce generalized adenopathy in adults.

The site of localized or regional adenopathy may provide a useful clue about the cause. Occipital adenopathy often reflects an infection of the scalp, and preauricular adenopathy accompanies conjunctival infections and cat-scratch disease. The most frequent site of regional adenopathy is the neck, and most of the causes are benign—upper respiratory infections, oral and dental lesions, infectious mononucleosis, or other viral illnesses. The chief malignant causes include metastatic cancer from head and neck, breast, lung, and thyroid primaries. Enlargement of supraclavicular and scalene nodes is always abnormal. Because these nodes drain regions of the lung and retroperitoneal space, they can reflect lymphomas, other cancers, or infectious processes arising in these areas. Virchow's node is an enlarged left supraclavicular node infiltrated with metastatic cancer from a gastrointestinal primary. Metastases to supraclavicular nodes also occur from lung, breast, testis, or ovarian cancers. Tuberculosis, sarcoidosis, and toxoplasmosis are nonneoplastic causes of supraclavicular adenopathy. Axillary adenopathy is usually due to injuries or localized infections of the ipsilateral upper extremity. Malignant causes include melanoma or lymphoma and, in women, breast cancer. Inguinal lymphadenopathy is usually secondary to infections or trauma of the lower extremities and may accompany sexually transmitted diseases such as lymphogranuloma venereum, primary syphilis, genital herpes, or chancroid. These nodes may also be involved by lymphomas and metastatic cancer from primary lesions of the rectum, genitalia, or lower extremities (melanoma).

The size and texture of the lymph node(s) and the presence of pain are useful parameters in evaluating a patient with lymphadenopathy. Nodes $<1.0 \text{ cm}^2$ in area ($1.0 \text{ cm} \times 1.0 \text{ cm}$ or less) are almost always secondary to benign, nonspecific reactive causes. In one retrospective analysis of younger patients (9–25 years) who had a lymph node biopsy, a maximum diameter of $>2 \text{ cm}$ served as one discriminant for predicting that the biopsy would reveal malignant or granulomatous disease. Another study showed that a lymph node size of 2.25 cm^2 ($1.5 \text{ cm} \times 1.5 \text{ cm}$) was the best size limit for distinguishing malignant or granulomatous lymphadenopathy from other causes of lymphadenopathy. Patients with node(s) $\leq 1.0 \text{ cm}^2$ should be observed after excluding infectious mononucleosis and/or toxoplasmosis unless there are symptoms and signs of an underlying systemic illness.

The texture of lymph nodes may be described as soft, firm, rubbery, hard, discrete, matted, tender, movable, or fixed. Tenderness is found when the capsule is stretched during rapid enlargement, usually secondary to an inflammatory process. Some malignant diseases such as acute leukemia may produce rapid enlargement and pain in the nodes. Nodes involved by lymphoma tend to be large, discrete, symmetric, rubbery, firm, mobile, and nontender. Nodes containing metastatic cancer are often hard, nontender, and nonmovable because of fixation to surrounding tissues. The coexistence of splenomegaly in the patient with lymphadenopathy implies a systemic illness such

as infectious mononucleosis, lymphoma, acute or chronic leukemia, SLE, sarcoidosis, toxoplasmosis, cat-scratch disease, or other less common hematologic disorders. The patient's story should provide helpful clues about the underlying systemic illness.

Nonsuperficial presentations (thoracic or abdominal) of adenopathy are usually detected as the result of a symptom-directed diagnostic workup. Thoracic adenopathy may be detected by routine chest radiography or during the workup for superficial adenopathy. It may also be found because the patient complains of a cough or wheezing from airway compression; hoarseness from recurrent laryngeal nerve involvement; dysphagia from esophageal compression; or swelling of the neck, face, or arms secondary to compression of the superior vena cava or subclavian vein. The differential diagnosis of mediastinal and hilar adenopathy includes primary lung disorders and systemic illnesses that characteristically involve mediastinal or hilar nodes. In the young, mediastinal adenopathy is associated with infectious mononucleosis and sarcoidosis. In endemic regions, histoplasmosis can cause unilateral paratracheal lymph node involvement that mimics lymphoma. Tuberculosis can also cause unilateral adenopathy. In older patients, the differential diagnosis includes primary lung cancer (especially among smokers), lymphomas, metastatic carcinoma (usually lung), tuberculosis, fungal infection, and sarcoidosis.

Enlarged intraabdominal or retroperitoneal nodes are usually malignant. Although tuberculosis may present as mesenteric lymphadenitis, these masses usually contain lymphomas or, in young men, germ cell tumors.

LABORATORY INVESTIGATION

The laboratory investigation of patients with lymphadenopathy must be tailored to elucidate the etiology suspected from the patient's history and physical findings. One study from a family practice clinic evaluated 249 younger patients with "enlarged lymph nodes, not infected" or "lymphadenitis." No laboratory studies were obtained in 51%. When studies were performed, the most common were a complete blood count (CBC) (33%), throat culture (16%), chest x-ray (12%), or monospot test (10%). Only eight patients (3%) had a node biopsy, and half of those were normal or reactive. The CBC can provide useful data for the diagnosis of acute or chronic leukemias, EBV or CMV mononucleosis, lymphoma with a leukemic component, pyogenic infections, or immune cytopenias in illnesses such as SLE. Serologic studies may demonstrate antibodies specific to components of EBV, CMV, HIV, and other viruses; *Toxoplasma gondii*; *Brucella*; etc. If SLE is suspected, antinuclear and anti-DNA antibody studies are warranted.

The chest x-ray is usually negative, but the presence of a pulmonary infiltrate or mediastinal lymphadenopathy would suggest tuberculosis, histoplasmosis, sarcoidosis, lymphoma, primary lung cancer, or metastatic cancer and demands further investigation.

A variety of imaging techniques (CT, MRI, ultrasound, color Doppler ultrasonography) have been employed to differentiate benign from malignant lymph nodes, especially in patients with head and neck cancer. CT and MRI are comparably accurate (65–90%) in the diagnosis of metastases to cervical lymph nodes. Ultrasonography has been used to determine the long (L) axis, short (S) axis, and a ratio of long to short axis in cervical nodes. An L/S ratio of <2.0 has a sensitivity and a specificity of 95% for distinguishing benign and malignant nodes in patients with head and neck cancer. This ratio has greater specificity and sensitivity than palpation or measurement of either the long or the short axis alone.

The indications for lymph node biopsy are imprecise, yet it is a valuable diagnostic tool. The decision to biopsy may be made early in a patient's evaluation or delayed for up to 2 weeks. Prompt biopsy should occur if the patient's history and physical findings suggest a malignancy; examples include a solitary, hard, nontender cervical node in an older patient who is a chronic user of tobacco; supraclavicular adenopathy; and solitary or generalized adenopathy that is firm, movable, and suggestive of lymphoma. If a primary head

and neck cancer is suspected as the basis of a solitary, hard cervical node, then a careful ENT examination should be performed. Any mucosal lesion that is suspicious for a primary neoplastic process should be biopsied first. If no mucosal lesion is detected, an excisional biopsy of the largest node should be performed. Fine-needle aspiration should not be performed as the first diagnostic procedure. Most diagnoses require more tissue than such aspiration can provide, and it often delays a definitive diagnosis. Fine-needle aspiration should be reserved for thyroid nodules and for confirmation of relapse in patients whose primary diagnosis is known. If the primary physician is uncertain about whether to proceed to biopsy, consultation with a hematologist or medical oncologist should be helpful. In primary care practices, <5% of lymphadenopathy patients will require a biopsy. That percentage will be considerably larger in referral practices, i.e., hematology, oncology, or ENT.

Two groups have reported algorithms that they claim will identify more precisely those lymphadenopathy patients who should have a biopsy. Both reports were retrospective analyses in referral practices. The first study involved patients 9–25 years of age who had a node biopsy performed. Three variables were identified that predicted those young patients with peripheral lymphadenopathy who should undergo biopsy; lymph node size >2 cm in diameter and abnormal chest x-ray had positive predictive values, whereas recent ENT symptoms had negative predictive values. The second study evaluated 220 lymphadenopathy patients in a hematology unit and identified five variables (lymph node size, location [supraclavicular or nonsupraclavicular], age [>40 years or <40 years], texture [nonhard or hard], and tenderness) that were used in a mathematical model to identify those patients requiring a biopsy. Positive predictive value was found for age >40 years, supraclavicular location, node size >2.25 cm², hard texture, and lack of pain or tenderness. Negative predictive value was evident for age <40 years, node size <1.0 cm², nonhard texture, and tender or painful nodes. Ninety-one percent of those who required biopsy were correctly classified by this model. Because both of these studies were retrospective analyses and one was limited to young patients, it is not known how useful these models would be if applied prospectively in a primary care setting.

Most lymphadenopathy patients do not require a biopsy, and at least half require no laboratory studies. If the patient's history and physical findings point to a benign cause for lymphadenopathy, careful follow-up at a 2- to 4-week interval can be employed. The patient should be instructed to return for reevaluation if there is an increase in the size of the nodes. Antibiotics are not indicated for lymphadenopathy unless strong evidence of a bacterial infection is present. Glucocorticoids should not be used to treat lymphadenopathy because their lympholytic effect obscures some diagnoses (lymphoma, leukemia, Castleman's disease) and they contribute to delayed healing or activation of underlying infections. An exception to this statement is the life-threatening pharyngeal obstruction by enlarged lymphoid tissue in Waldeyer's ring that is occasionally seen in infectious mononucleosis.

SPLENOMEGLY

■ STRUCTURE AND FUNCTION OF THE SPLEEN

The spleen is a reticuloendothelial organ that has its embryologic origin in the dorsal mesogastrium at about 5 weeks' gestation. It arises in a series of hillocks, migrates to its normal adult location in the left upper quadrant (LUQ), and is attached to the stomach via the gastrosplenic ligament and to the kidney via the lienorenal ligament. When the hillocks fail to unify into a single tissue mass, accessory spleens may develop in around 20% of persons. The function of the spleen has been elusive. Galen believed it was the source of "black bile" or melancholia, and the word *hypochondria* (literally, beneath the ribs) and the idiom "to vent one's spleen" attest to the beliefs that the spleen had an important influence on the psyche and emotions. In humans, its normal physiologic roles seem to be the following:

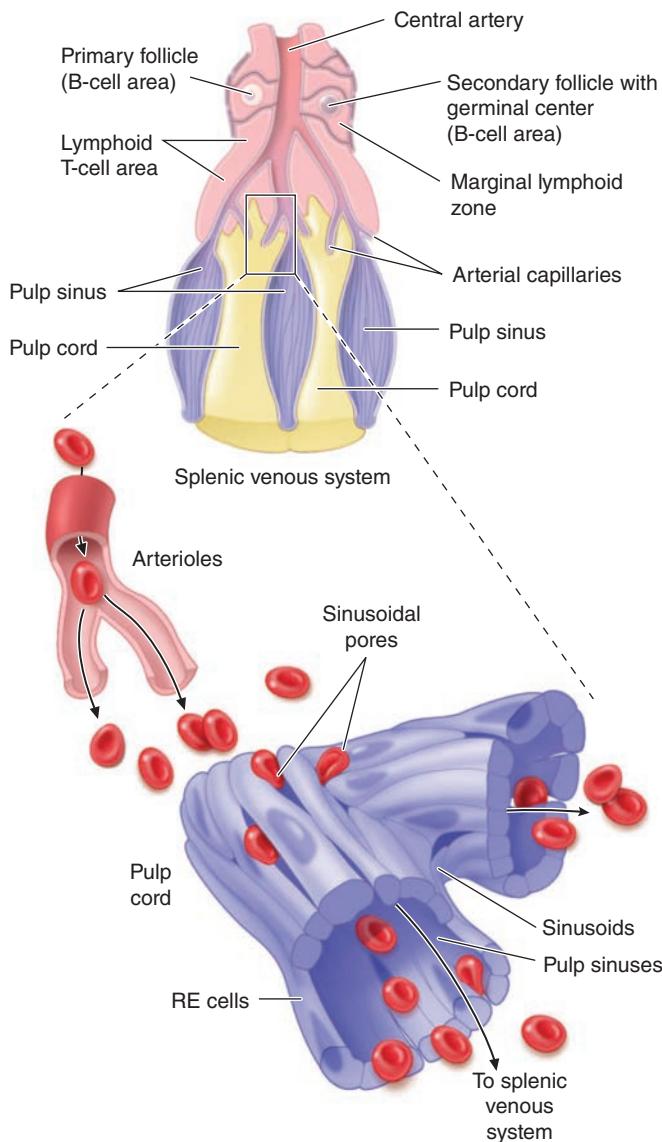


FIGURE 66-1 Schematic spleen structure. The spleen comprises many units of red and white pulp centered around small branches of the splenic artery, called *central arteries*. White pulp is lymphoid in nature and contains B-cell follicles, a marginal zone around the follicles, and T-cell-rich areas sheathing arterioles. The red pulp areas include pulp sinuses and pulp cords. The cords are dead ends. In order to regain access to the circulation, red blood cells must traverse tiny openings in the sinusoidal lining. Stiff, damaged, or old red cells cannot enter the sinuses. RE, reticuloendothelial. (Bottom portion of figure reproduced with permission from RS Hillman, KA Ault: *Hematology in Clinical Practice*, 4th ed. New York, McGraw-Hill, 2005.)

1. Maintenance of quality control over erythrocytes in the red pulp by removal of senescent and defective red blood cells. The spleen accomplishes this function through a unique organization of its parenchyma and vasculature (Fig. 66-1).
2. Synthesis of antibodies in the white pulp.
3. The removal of antibody-coated bacteria and antibody-coated blood cells from the circulation.

An increase in these normal functions may result in splenomegaly.

The spleen is composed of *red pulp* and *white pulp*, which are Malpighi's terms for the red blood-filled sinuses and reticuloendothelial cell-lined cords and the white lymphoid follicles arrayed within the red pulp matrix. The spleen is in the portal circulation. The reason for this is unknown but may relate to the fact that lower blood pressure allows less rapid flow and minimizes damage to normal erythrocytes. Blood flows into the spleen at a rate of about 150 mL/min through the splenic artery, which ultimately ramifies into central arterioles. Some blood goes from the arterioles to capillaries and then to splenic veins

and out of the spleen, but the majority of blood from central arterioles flows into the macrophage-lined sinuses and cords. The blood entering the sinuses reenters the circulation through the splenic venules, but the blood entering the cords is subjected to an inspection of sorts. To return to the circulation, the blood cells in the cords must squeeze through slits in the cord lining to enter the sinuses that lead to the venules. Old and damaged erythrocytes are less deformable and are retained in the cords, where they are destroyed and their components recycled. Red cell-inclusion bodies such as parasites (Chaps. 224, 225, and A2), nuclear residua (Howell-Jolly bodies, see Fig. 63-6), or denatured hemoglobin (Heinz bodies) are pinched off in the process of passing through the slits, a process called *pitting*. The culling of dead and damaged cells and the pitting of cells with inclusions appear to occur without significant delay because the blood transit time through the spleen is only slightly slower than in other organs.

The spleen is also capable of assisting the host in adapting to its hostile environment. It has at least three adaptive functions: (1) clearance of bacteria and particulates from the blood, (2) the generation of immune responses to certain pathogens, and (3) the generation of cellular components of the blood under circumstances in which the marrow is unable to meet the needs (i.e., extramedullary hematopoiesis). The latter adaptation is a recapitulation of the blood-forming function the spleen plays during gestation. In some animals, the spleen also serves a role in the vascular adaptation to stress because it stores red blood cells (often hemoconcentrated to higher hematocrits than normal) under normal circumstances and contracts under the influence of β -adrenergic stimulation to provide the animal with an autotransfusion and improved oxygen-carrying capacity. However, the normal human spleen does not sequester or store red blood cells and does not contract in response to sympathetic stimuli. The normal human spleen contains approximately one-third of the total body platelets and a significant number of marginated neutrophils. These sequestered cells are available when needed to respond to bleeding or infection.

APPROACH TO THE PATIENT

Splenomegaly

CLINICAL ASSESSMENT

The most common *symptoms* produced by diseases involving the spleen are pain and a heavy sensation in the LUQ. Massive splenomegaly may cause early satiety. Pain may result from acute swelling of the spleen with stretching of the capsule, infarction, or inflammation of the capsule. For many years, it was believed that splenic infarction was clinically silent, which, at times, is true. However, Soma Weiss, in his classic 1942 report of the self-observations by a Harvard medical student on the clinical course of subacute bacterial endocarditis, documented that severe LUQ and pleuritic chest pain may accompany thromboembolic occlusion of splenic blood flow. Vascular occlusion, with infarction and pain, is commonly seen in children with sickle cell crises. Rupture of the spleen, from either trauma or infiltrative disease that breaks the capsule, may result in intraperitoneal bleeding, shock, and death. The rupture itself may be painless.

A palpable spleen is the major *physical sign* produced by diseases affecting the spleen and suggests enlargement of the organ. The normal spleen weighs <250 g, decreases in size with age, normally lies entirely within the rib cage, has a maximum cephalocaudad diameter of 13 cm by ultrasonography or maximum length of 12 cm and/or width of 7 cm by radionuclide scan, and is usually not palpable. However, a palpable spleen was found in 3% of 2200 asymptomatic, male, freshman college students. Follow-up at 3 years revealed that 30% of those students still had a palpable spleen without any increase in disease prevalence. Ten-year follow-up found no evidence for lymphoid malignancies. Furthermore, in some tropical countries (e.g., New Guinea), the incidence of splenomegaly may reach 60%. Thus, the presence of a palpable spleen does not always equate with presence of disease. Even when disease is present,

splenomegaly may not reflect the primary disease but rather a reaction to it. For example, in patients with Hodgkin's disease, only two-thirds of the palpable spleens show involvement by the cancer.

Physical examination of the spleen uses primarily the techniques of palpation and percussion. Inspection may reveal fullness in the LUQ that descends on inspiration, a finding associated with a massively enlarged spleen. Auscultation may reveal a venous hum or friction rub.

Palpation can be accomplished by bimanual palpation, ballotment, and palpation from above (Middleton maneuver). For bimanual palpation, which is at least as reliable as the other techniques, the patient is supine with flexed knees. The examiner's left hand is placed on the lower rib cage and pulls the skin toward the costal margin, allowing the fingertips of the right hand to feel the tip of the spleen as it descends while the patient inspires slowly, smoothly, and deeply. Palpation is begun with the right hand in the left lower quadrant with gradual movement toward the left costal margin, thereby identifying the lower edge of a massively enlarged spleen. When the spleen tip is felt, the finding is recorded as centimeters below the left costal margin at some arbitrary point, i.e., 10–15 cm, from the midpoint of the umbilicus or the xiphisternal junction. This allows other examiners to compare findings or the initial examiner to determine changes in size over time. Bimanual palpation in the right lateral decubitus position adds nothing to the supine examination.

Percussion for splenic dullness is accomplished with any of three techniques described by Nixon, Castell, or Barkun:

1. *Nixon's method:* The patient is placed on the right side so that the spleen lies above the colon and stomach. Percussion begins at the lower level of pulmonary resonance in the posterior axillary line and proceeds diagonally along a perpendicular line toward the lower midanterior costal margin. The upper border of dullness is normally 6–8 cm above the costal margin. Dullness >8 cm in an adult is presumed to indicate splenic enlargement.
2. *Castell's method:* With the patient supine, percussion in the lowest intercostal space in the anterior axillary line (8th or 9th) produces a resonant note if the spleen is normal in size. This is true during expiration or full inspiration. A dull percussion note on full inspiration suggests splenomegaly.
3. *Percussion of Traube's semilunar space:* The borders of Traube's space are the sixth rib superiorly, the left midaxillary line laterally, and the left costal margin inferiorly. The patient is supine with the left arm slightly abducted. During normal breathing, this space is percussed from medial to lateral margins, yielding a normal resonant sound. A dull percussion note suggests splenomegaly.

Studies comparing methods of percussion and palpation with a standard of ultrasonography or scintigraphy have revealed sensitivity of 56–71% for palpation and 59–82% for percussion. Reproducibility among examiners is better for palpation than percussion. Both techniques are less reliable in obese patients or patients who have just eaten. Thus, the physical examination techniques of palpation and percussion are imprecise at best. It has been suggested that the examiner perform percussion first and, if positive, proceed to palpation; if the spleen is palpable, then one can be reasonably confident that splenomegaly exists. However, not all LUQ masses are enlarged spleens; gastric or colon tumors and pancreatic or renal cysts or tumors can mimic splenomegaly.

The presence of an enlarged spleen can be more precisely determined, if necessary, by liver-spleen radionuclide scan, CT, MRI, or ultrasonography. The latter technique is the current procedure of choice for routine assessment of spleen size (normal = a maximum cephalocaudad diameter of 13 cm) because it has high sensitivity and specificity and is safe, noninvasive, quick, mobile, and less costly. Equipment advances allow ultrasonography to be performed at the bedside with excellent sensitivity and specificity. Nuclear medicine scans are accurate, sensitive, and reliable but are costly,

require greater time to generate data, and use immobile equipment. They have the advantage of demonstrating accessory splenic tissue. CT and MRI provide accurate determination of spleen size, but the equipment is immobile and the procedures are expensive. MRI appears to offer no advantage over CT. Changes in spleen structure such as mass lesions, infarcts, inhomogeneous infiltrates, and cysts are more readily assessed by CT, MRI, or ultrasonography. None of these techniques is very reliable in the detection of patchy infiltration (e.g., Hodgkin's disease).

DIFFERENTIAL DIAGNOSIS

Many of the diseases associated with splenomegaly are listed in **Table 66-2**. They are grouped according to the presumed basic mechanisms responsible for organ enlargement:

1. Hyperplasia or hypertrophy related to a particular splenic function such as reticuloendothelial hyperplasia (work hypertrophy) in diseases such as hereditary spherocytosis or thalassemia syndromes that require removal of large numbers of defective red blood cells; immune hyperplasia in response to systemic infection (infectious mononucleosis, subacute bacterial endocarditis) or to immunologic diseases (immune thrombocytopenia, SLE, Felty's syndrome).
2. Passive congestion due to decreased blood flow from the spleen in conditions that produce portal hypertension (cirrhosis, Budd-Chiari syndrome, congestive heart failure).
3. Infiltrative diseases of the spleen (lymphomas, metastatic cancer, amyloidosis, Gaucher's disease, myeloproliferative disorders with extramedullary hematopoiesis).

The differential diagnostic possibilities are much fewer when the spleen is "massively enlarged" or palpable >8 cm below the left costal margin or its drained weight is ≥ 1000 g (**Table 66-3**). The vast majority of such patients will have non-Hodgkin's lymphoma, chronic lymphocytic leukemia, hairy cell leukemia, chronic myeloid leukemia, myelofibrosis with myeloid metaplasia, or polycythemia vera.

LABORATORY ASSESSMENT

The major laboratory abnormalities accompanying splenomegaly are determined by the underlying systemic illness. Erythrocyte counts may be normal, decreased (thalassemia major syndromes, SLE, cirrhosis with portal hypertension), or increased (polycythemia vera). Granulocyte counts may be normal, decreased (Felty's syndrome, congestive splenomegaly, leukemias), or increased (infections or inflammatory disease, myeloproliferative disorders). Similarly, the platelet count may be normal, decreased when there is enhanced sequestration or destruction of platelets in an enlarged spleen (congestive splenomegaly, Gaucher's disease, immune thrombocytopenia), or increased in the myeloproliferative disorders such as polycythemia vera.

The CBC may reveal cytopenia of one or more blood cell types, which should suggest *hypersplenism*. This condition is characterized by splenomegaly, cytopenia(s), normal or hyperplastic bone marrow, and a response to splenectomy. The latter characteristic is less precise because reversal of cytopenia, particularly granulocytopenia, is sometimes not sustained after splenectomy. The cytopenias result from increased destruction of the cellular elements secondary to reduced flow of blood through enlarged and congested cords (congestive splenomegaly) or to immune-mediated mechanisms. In hypersplenism, various cell types usually have normal morphology on the peripheral blood smear, although the red cells may be spherocytic due to loss of surface area during their longer transit through the enlarged spleen. The increased marrow production of red cells should be reflected as an increased reticulocyte production index, although the value may be less than expected due to increased sequestration of reticulocytes in the spleen.

The need for additional laboratory studies is dictated by the differential diagnosis of the underlying illness of which splenomegaly is a manifestation.

TABLE 66-2 Diseases Associated with Splenomegaly Grouped by Pathogenic Mechanism**Enlargement Due to Increased Demand for Splenic Function**

Reticuloendothelial system hyperplasia (for removal of defective erythrocytes)	Leishmaniasis
Spherocytosis	Trypanosomiasis
Early sickle cell anemia	Ehrlichiosis
Ovalocytosis	Disordered immunoregulation
Thalassemia major	Hemophagocytic lymphohistiocytosis (HLH)
Hemoglobinopathies	Rheumatoid arthritis (Felty's syndrome)
Paroxysmal nocturnal hemoglobinuria	Systemic lupus erythematosus
Pernicious anemia	Collagen vascular diseases
Immune hyperplasia	Serum sickness
Response to infection (viral, bacterial, fungal, parasitic)	Immune hemolytic anemias
Infectious mononucleosis	Immune thrombocytopenias
AIDS	Immune neutropenias
Viral hepatitis	Drug reactions
Cytomegalovirus	Angioimmunoblastic lymphadenopathy
Subacute bacterial endocarditis	Sarcoidosis
Bacterial septicemia	Thyrotoxicosis (benign lymphoid hypertrophy)
Congenital syphilis	Interleukin 2 therapy
Splenic abscess	Extramedullary hematopoiesis
Tuberculosis	Myelofibrosis
Histoplasmosis	Marrow damage by toxins, radiation, strontium
Malaria	Marrow infiltration by tumors, leukemias, Gaucher's disease

Enlargement Due to Abnormal Splenic or Portal Blood Flow

Cirrhosis	Splenic artery aneurysm
Hepatic vein obstruction	Hepatic schistosomiasis
Portal vein obstruction, intrahepatic or extrahepatic	Congestive heart failure
Cavernous transformation of the portal vein	Hepatic echinococcosis
Splenic vein obstruction	Portal hypertension (any cause including the above): "Banti's disease"

Infiltration of the Spleen

Intracellular or extracellular depositions	Hodgkin's disease
Amyloidosis	Myeloproliferative syndromes (e.g., polycythemia vera, essential thrombocythosis)
Gaucher's disease	Angiosarcomas
Niemann-Pick disease	Metastatic tumors (melanoma is most common)
Tangier disease	Eosinophilic granuloma
Hurler's syndrome and other mucopolysaccharidoses	Histiocytosis X
Hyperlipidemias	Hamartomas
Benign and malignant cellular infiltrations	Hemangiomas, fibromas, lymphangiomas
Leukemias (acute, chronic, lymphoid, myeloid, monocytic)	Splenic cysts
Lymphomas	

Unknown Etiology

Idiopathic splenomegaly	Iron-deficiency anemia
Berylliosis	

SPLENECTOMY

Splenectomy is infrequently performed for diagnostic purposes, especially in the absence of clinical illness or other diagnostic tests that suggest underlying disease. More often, splenectomy is performed for symptom control in patients with massive splenomegaly, for disease

control in patients with traumatic splenic rupture, or for correction of cytopenias in patients with hypersplenism or immune-mediated destruction of one or more cellular blood elements. Splenectomy is necessary for staging of patients with Hodgkin's disease only in those with clinical stage I or II disease in whom radiation therapy alone is contemplated as the treatment. Noninvasive staging of the spleen in Hodgkin's disease is not a sufficiently reliable basis for treatment decisions because one-third of normal-sized spleens will be involved with Hodgkin's disease and one-third of enlarged spleens will be tumor-free. The widespread use of systemic therapy to treat all stages of Hodgkin's disease has made staging laparotomy with splenectomy unnecessary. Although splenectomy in chronic myeloid leukemia (CML) does not affect the natural history of disease, removal of the massive spleen usually makes patients significantly more comfortable and simplifies their management by significantly reducing transfusion requirements.

TABLE 66-3 Diseases Associated with Massive Splenomegaly*

Chronic myeloid leukemia	Gaucher's disease
Lymphomas	Chronic lymphocytic leukemia
Hairy cell leukemia	Sarcoidosis
Myelofibrosis with myeloid metaplasia	Autoimmune hemolytic anemia
Polycythemia vera	Diffuse splenic hemangiomatosis

*The spleen extends >8 cm below left costal margin and/or weighs >1000 g.

The improvements in therapy of CML have reduced the need for splenectomy for symptom control. Splenectomy is an effective secondary or tertiary treatment for two chronic B-cell leukemias, hairy cell leukemia and prolymphocytic leukemia, and for the very rare splenic mantle cell or marginal zone lymphoma. Splenectomy in these diseases may be associated with significant tumor regression in bone marrow and other sites of disease. Similar regressions of systemic disease have been noted after splenic irradiation in some types of lymphoid tumors, especially chronic lymphocytic leukemia and prolymphocytic leukemia. This has been termed the *abscopal effect*. Such systemic tumor responses to local therapy directed at the spleen suggest that some hormone or growth factor produced by the spleen may affect tumor cell proliferation, but this conjecture is not yet substantiated. A common therapeutic indication for splenectomy is traumatic or iatrogenic splenic rupture. In a fraction of patients with splenic rupture, peritoneal seeding of splenic fragments can lead to *splenosis*—the presence of multiple rests of spleen tissue not connected to the portal circulation. This ectopic spleen tissue may cause pain or gastrointestinal obstruction, as in endometriosis. A large number of hematologic, immunologic, and congestive causes of splenomegaly can lead to destruction of one or more cellular blood elements. In the majority of such cases, splenectomy can correct the cytopenias, particularly anemia and thrombocytopenia. In a large series of patients seen in two tertiary care centers, the indication for splenectomy was diagnostic in 10% of patients, therapeutic in 44%, staging for Hodgkin's disease in 20%, and incidental to another procedure in 26%. Perhaps the only contraindication to splenectomy is the presence of marrow failure, in which the enlarged spleen is the only source of hematopoietic tissue.

Often the splenectomy is done by laparoscopy, which is associated with shorter hospital stays and faster recovery than the open procedure; however, concern has emerged that the laparoscopic approach is associated with a higher risk of postoperative portal venous system thrombosis and Budd-Chiari syndrome.

The absence of the spleen has minimal long-term effects on the hematologic profile. In the immediate postsplenectomy period, leukocytosis (up to 25,000/ μ L) and thrombocytosis (up to $1 \times 10^6/\mu$ L) may develop, but within 2–3 weeks, blood cell counts and survival of each cell lineage are usually normal. The chronic manifestations of splenectomy are marked variation in size and shape of erythrocytes (anisocytosis, poikilocytosis) and the presence of Howell-Jolly bodies (nuclear remnants), Heinz bodies (denatured hemoglobin), basophilic stippling, and an occasional nucleated erythrocyte in the peripheral blood. When such erythrocyte abnormalities appear in a patient whose spleen has not been removed, one should suspect splenic infiltration by tumor that has interfered with its normal culling and pitting function.

The most serious consequence of splenectomy is increased susceptibility to bacterial infections, particularly those with capsules such as *Streptococcus pneumoniae*, *Haemophilus influenzae*, and some gram-negative enteric organisms. Patients aged <20 years are particularly susceptible to overwhelming sepsis with *S. pneumoniae*, and the overall actuarial risk of sepsis in patients who have had their spleens removed is about 7% in 10 years. The case-fatality rate for pneumococcal sepsis in splenectomized patients is 50–80%. About 25% of patients without spleens will develop a serious infection at some time in their life. The frequency is highest within the first 3 years after splenectomy. About 15% of the infections are polymicrobial, and lung, skin, and blood are the most common sites. No increased risk of viral infection has been noted in patients who have no spleen. The susceptibility to bacterial infections relates to the inability to remove opsonized bacteria from the bloodstream and a defect in making antibodies to T-cell-independent antigens such as the polysaccharide components of bacterial capsules. Pneumococcal vaccine should be administered to all patients 2 weeks before elective splenectomy. The Advisory Committee on Immunization Practices recommends that these patients receive

repeat vaccination 5 years after splenectomy. Efficacy has not been proven for this group, and the recommendation discounts the possibility that administration of the vaccine may actually lower the titer of specific pneumococcal antibodies. A more effective pneumococcal conjugate vaccine that involves T cells in the response is now available (PCV13). The vaccine to *Neisseria meningitidis* should also be given to patients in whom elective splenectomy is planned. Although efficacy data for *Haemophilus influenzae* type b vaccine are not available for older children or adults, it may be given to patients who have had a splenectomy.

Splenectomized patients should be educated to consider any unexplained fever as a medical emergency. Prompt medical attention with evaluation and treatment of suspected bacteremia may be lifesaving. Routine chemoprophylaxis with oral penicillin can result in the emergence of drug-resistant strains and is not recommended.

In addition to an increased susceptibility to bacterial infections, splenectomized patients are also more susceptible to the parasitic disease babesiosis. The splenectomized patient should avoid areas where the parasite *Babesia* is endemic (e.g., Cape Cod, MA).

Surgical removal of the spleen is an obvious cause of hyposplenism. Patients with sickle cell disease often suffer from autosplenectomy as a result of splenic destruction by the numerous infarcts associated with sickle cell crises during childhood. Indeed, the presence of a palpable spleen in a patient with sickle cell disease after age 5 suggests a coexisting hemoglobinopathy, e.g., thalassemia or hemoglobin C. In addition, patients who receive splenic irradiation for a neoplastic or autoimmune disease are also functionally hyposplenic. The term *hyposplenism* is preferred to *asplenism* in referring to the physiologic consequences of splenectomy because asplenia is a rare, specific, and fatal congenital abnormality in which there is a failure of the left side of the coelomic cavity (which includes the splenic anlagen) to develop normally. Infants with asplenia have no spleens, but that is the least of their problems. The right side of the developing embryo is duplicated on the left so there is liver where the spleen should be, there are two right lungs, and the heart comprises two right atria and two right ventricles.

ACKNOWLEDGMENT

Patrick H. Henry, MD, friend and mentor now deceased, contributed significantly to the chapter in past editions, and much of his work remains in this chapter.

■ FURTHER READING

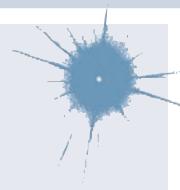
- BARKUN AN et al: The bedside assessment of splenic enlargement. *Am J Med* 91:512, 1991.
- CESSFORD T et al: Comparing physical examination with sonographic versions of the same examination techniques for splenomegaly. *J Ultrasound Med* 37:1621, 2018.
- FACCHETTI F: Tumors of the spleen. *Int J Surg Pathol* 18:136S, 2010.
- GIRARD E et al: Management of splenic and pancreatic trauma. *J Visc Surg* 153(suppl 4):45, 2016.
- GRAVES SA et al: Does this patient have splenomegaly? *JAMA* 270:2218, 1993.
- KIM DK et al: Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older—United States, 2017. *MMWR* 66:136, 2017.
- KRAUS MD et al: The spleen as a diagnostic specimen: A review of ten years' experience at two tertiary care institutions. *Cancer* 91:2001, 2001.
- MCINTYRE OR, EBAUGH FG JR: Palpable spleens: Ten-year follow-up. *Ann Intern Med* 90:130, 1979.
- PANGALIS GA et al: Clinical approach to lymphadenopathy. *Semin Oncol* 20:570, 1993.
- WILLIAMSON HA JR: Lymphadenopathy in a family practice: A descriptive study of 240 cases. *J Fam Pract* 20:449, 1985.

This page intentionally left blank

67

Principles of Clinical Pharmacology

Dan M. Roden



Drugs are the cornerstone of modern therapeutics. Nevertheless, it is well recognized among health care providers and the lay community that the outcome of drug therapy varies widely among individuals. While this variability has been perceived as an unpredictable, and therefore inevitable, accompaniment of drug therapy, this is not the case.

Drugs interact with specific target molecules to produce their beneficial and adverse effects. The chain of events between administration of a drug and production of these effects in the body can be divided into two components, both of which contribute to variability in drug actions. The first component comprises the processes that determine drug delivery to, and removal from, molecular targets. The resulting description of the relationship between drug concentration and time is termed *pharmacokinetics*. The second component of variability in drug action comprises the processes that determine variability in drug actions independent of variability in drug delivery to effector drug sites. This description of the relationship between drug concentration and effect is termed *pharmacodynamics*. As discussed further below, pharmacodynamic variability can arise as a result of variability in function of the target molecule itself or of variability in the broad biologic context in which the drug-target interaction occurs to achieve drug effects. The principles described below were developed by studying small drug molecules but are equally useful in describing the effects of very large molecules, such as the therapeutic antibodies increasingly applied to autoimmune diseases and cancer.

Two important goals of clinical pharmacology are (1) to provide a description of conditions under which drug actions vary among human subjects; and (2) to determine mechanisms underlying this variability, with the goal of improving therapy with available drugs as well as pointing to mechanisms whose targeting by new drugs may be effective in the treatment of human disease. The drug development process is briefly described at the end of this chapter.

The first steps in the discipline of clinical pharmacology were empirical descriptions of the influence of disease on drug actions and of individuals or families with unusual sensitivities to adverse drug reactions (ADRs). These important descriptive findings are now being replaced by an understanding of the molecular mechanisms underlying variability in drug actions. Importantly, it is often the personal interaction of the patient with the physician or other health care provider that first identifies unusual variability in drug actions; maintained alertness to unusual drug responses continues to be a key component of improving drug safety.

One useful unifying framework is to consider that the effects of disease, drug coadministration, or familial factors in modulating drug action reflect variability in expression or function of specific genes whose products determine pharmacokinetics and pharmacodynamics. This idea forms the basis for pharmacogenomic science; a few examples are cited in this chapter, and further details are addressed in [Chap. 68](#).

■ GLOBAL CONSIDERATIONS

It is true across all cultures and diseases that factors such as compliance, genetic variants affecting pharmacokinetics or pharmacodynamics (which themselves vary by ancestry), and drug interactions contribute to drug responses. Cost issues or cultural factors may determine the likelihood that specific drugs, drug combinations, or over-the-counter (OTC) remedies are prescribed. The broad principles of clinical pharmacology enunciated here can be used to analyze the mechanisms underlying successful or unsuccessful therapy with any drug.

■ INDICATIONS FOR DRUG THERAPY: RISK VERSUS BENEFIT

It is self-evident that the benefits of drug therapy should outweigh the risks. Benefits fall into broad categories: alleviation of symptoms, prevention of disease progression or complications, and prolonged life. However, establishing the balance between risk and benefit for an individual patient is not always simple. In addition to variability seen even within highly controlled drug trials, patients treated in clinical settings may display responses that were not observed in trials, sometimes due to comorbidities that were trial exclusion criteria. In addition, therapies that provide symptomatic benefits but shorten life may be entertained in patients with serious and highly symptomatic diseases such as heart failure or cancer. These considerations illustrate the continuing, highly personal nature of the relationship between the prescriber and the patient.

Adverse Effects Some adverse effects are so common and so readily associated with drug therapy that they are identified very early during clinical use of a drug. By contrast, serious ADRs may be sufficiently uncommon that they escape detection for many years after a drug begins to be widely used. The issue of how to identify rare but serious ADRs (that can profoundly affect the benefit-risk perception in an individual patient) has not been satisfactorily resolved. Potential approaches range from an increased understanding of the molecular and genetic basis of variability in drug actions to expanded postmarketing surveillance mechanisms. None of these have been completely effective, so practitioners must be continuously vigilant to the possibility that unusual symptoms may be related to specific drugs, or combinations of drugs, that their patients receive.

Therapeutic Index Beneficial and adverse reactions to drug therapy can be described by a series of dose-response relations ([Fig. 67-1](#)). Well-tolerated drugs demonstrate a wide margin, termed the *therapeutic ratio*, *therapeutic index*, or *therapeutic window*, between the doses required to produce a therapeutic effect and those producing toxicity. In cases where there is a similar relationship between plasma drug concentration and effects, monitoring plasma concentrations can be a highly effective aid in managing drug therapy by enabling concentrations to be maintained above the minimum required to produce an effect and below the concentration range likely to produce toxicity. Such monitoring has been widely used to guide therapy with specific agents, such as certain antiarrhythmics, anticonvulsants, and antibiotics. Many of the principles in clinical pharmacology and

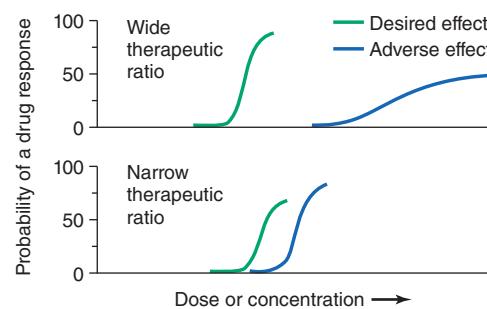


FIGURE 67-1 The concept of a therapeutic ratio. Each panel illustrates the relationship between increasing dose and cumulative probability of a desired or adverse drug effect. **Top.** A drug with a wide therapeutic ratio, that is, a wide separation of the two curves. **Bottom.** A drug with a narrow therapeutic ratio; here, the likelihood of adverse effects at therapeutic doses is increased because the curves are not well separated. Further, a steep dose-response curve for adverse effects is especially undesirable, as it implies that even small dosage increments may sharply increase the likelihood of toxicity. When there is a definable relationship between drug concentration (usually measured in plasma) and desirable and adverse effect curves, concentration may be substituted on the abscissa. Note that not all patients necessarily demonstrate a therapeutic response (or adverse effect) at any dose and that some effects (notably some adverse effects) may occur in a dose-independent fashion.

PRINCIPLES OF PHARMACOKINETICS

The processes of absorption, distribution, metabolism, and excretion—collectively termed *drug disposition*—determine the concentration of drug delivered to target effector molecules.

■ ABSORPTION AND BIOAVAILABILITY

When a drug is administered orally, subcutaneously, intramuscularly, rectally, sublingually, or directly into desired sites of action, the amount of drug eventually entering the systemic circulation may be less than with the intravenous route (Fig. 67-2A). The fraction of drug available to the systemic circulation by other routes is termed *bioavailability*. Bioavailability may be <100% for two main reasons: (1) incomplete absorption, or (2) metabolism or elimination prior to entering the systemic circulation.

Compared to the same dose given intravenously, a nonintravenous dose will have a later and lower peak plasma concentration (Fig. 67-2). Drug absorption may be reduced because a drug is incompletely released from its dosage form, undergoes destruction at the site of administration, or has physicochemical properties such as insolubility that prevent complete absorption from its site of administration. Slow absorption rates are deliberately designed into “slow-release” or “sustained-release” drug formulations in order to minimize variation in plasma concentrations during the interval between doses. Therapeutic antibodies administered subcutaneously may take days to reach the systemic circulation.

“First-Pass” Effect When a drug is administered orally, it must traverse the intestinal epithelium, the portal venous system, and the liver prior to entering the systemic circulation (Fig. 67-3). Once a drug enters the enterocyte, it may undergo metabolism, be transported into the portal vein, or be excreted back into the intestinal lumen. Both excretion into the intestinal lumen and metabolism decrease bioavailability. Once a drug passes this enterocyte barrier, it may also be taken up into the hepatocyte, where bioavailability can be further limited by metabolism or excretion into the bile. This elimination in

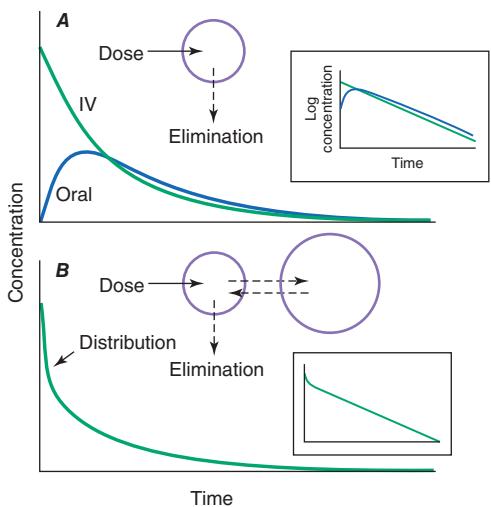


FIGURE 67-2 Idealized time-plasma concentration curves after a single dose of drug. **A.** The time course of drug concentration after an instantaneous intravenous (IV) bolus or an oral dose in the one-compartment model shown. The area under the time-concentration curve is clearly less with the oral drug than the IV drug, indicating incomplete bioavailability. Note that despite this incomplete bioavailability, concentration after the oral dose can be higher than after the IV dose at some time points. The inset shows that the decline of concentrations over time is linear on a log-linear plot, characteristic of first-order elimination, and that oral and IV drugs have the same elimination (parallel) time course. **B.** The decline of central compartment concentration when drug is distributed both to and from a peripheral compartment and eliminated from the central compartment. The rapid initial decline of concentration reflects not drug elimination but distribution.

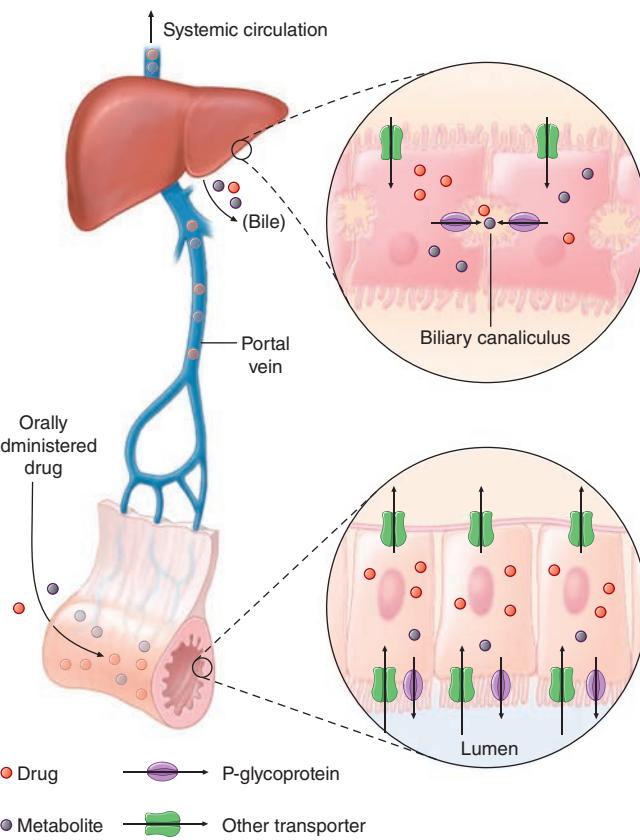


FIGURE 67-3 Mechanism of presystemic elimination. After drug enters the enterocyte, it can undergo metabolism, excretion into the intestinal lumen, or transport into the portal vein. Similarly, the hepatocyte may accomplish metabolism and biliary excretion prior to the entry of drug and metabolites to the systemic circulation. (Adapted by permission from DM Roden, in DP Zipes, J Jalife [eds]: Cardiac Electrophysiology: From Cell to Bedside, 4th ed. Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.)

intestine and liver, which reduces the amount of drug delivered to the systemic circulation, is termed *presystemic elimination*, *presystemic extraction*, or *first-pass elimination*.

■ DRUG TRANSPORT

Drug movement across the membrane of any cell, including enterocytes and hepatocytes, is a combination of passive diffusion and active transport, mediated by specific drug uptake and efflux molecules. One widely studied drug transport molecule is the drug efflux pump P-glycoprotein, the product of the *ABCB1* (or *MDR1*) gene. P-glycoprotein is expressed on the apical aspect of the enterocyte and on the canalicular aspect of the hepatocyte (Fig. 67-3). In both locations, it serves as an efflux pump, limiting availability of drug to the systemic circulation. P-glycoprotein-mediated drug efflux from cerebral capillaries limits drug brain penetration and is an important component of the blood-brain barrier. Other transporters mediate uptake into cells of drugs and endogenous substrates such as vitamins or nutrients.

■ DRUG METABOLISM

Drug metabolism generates compounds that are usually more polar and, hence, more readily excreted than parent drug. Metabolism takes place predominantly in the liver but can occur at other sites such as kidney, intestinal epithelium, lung, and plasma. Phase I metabolism involves chemical modification, most often oxidation accomplished by members of the cytochrome P450 (CYP) monooxygenase superfamily. CYPs and other molecules that are especially important for drug metabolism are presented in Table 67-1, and each drug may be a substrate for one or more of these enzymes. Phase II metabolism involves conjugation of specific endogenous compounds to drugs or their metabolites. The enzymes that accomplish phase II reactions include glucuronyl-, acetyl-, sulfo-, and methyltransferases. Drug metabolites

TABLE 67-1 Molecular Pathways Mediating Drug Disposition

ENZYME	SUBSTRATES ^a	INHIBITORS ^a
CYP3A	Calcium channel blockers Antiarrhythmics (lidocaine, quinidine, mexiletine) HMG-CoA reductase inhibitors ("statins"; see text) Cyclosporine, tacrolimus Indinavir, saquinavir, ritonavir	Amiodarone Ketoconazole, itraconazole Erythromycin, clarithromycin Ritonavir Gemfibrozil and other fibrates
CYP2D6 ^b	Timolol, metoprolol, carvedilol Propafenone, flecainide Tricyclic antidepressants Fluoxetine, paroxetine	Quinidine (even at ultra-low doses) Tricyclic antidepressants Fluoxetine, paroxetine
CYP2C9 ^b	Warfarin Phenytoin Glipizide Losartan	Amiodarone Fluconazole Phenytoin
CYP2C19 ^b	Omeprazole Mephenytoin Clopidogrel	Omeprazole
CYP2B6 ^b	Efavirenz	
Thiopurine S-methyltransferase ^b	6-Mercaptopurine, azathioprine	
N-acetyltransferase ^b	Isoniazid Procainamide Hydralazine Some sulfonamides	
UGT1A1 ^b	Irinotecan	
Pseudocholinesterase ^b	Succinylcholine	
TRANSPORTER	SUBSTRATES ^a	INHIBITORS ^a
P-glycoprotein	Digoxin HIV protease inhibitors Many CYP3A substrates	Quinidine Amiodarone Verapamil Cyclosporine Itraconazole Erythromycin
SLCO1B1 ^b	Simvastatin and some other statins	

^aInhibitors affect the molecular pathway and thus may decrease substrate metabolism. ^bClinically important genetic variants described; see Chap. 68.

Note: A listing of CYP substrates, inhibitors, and inducers is maintained at <https://drug-interactions.medicine.iu.edu/MainTable.aspx>.

may exert important pharmacologic activity, as discussed further below. Therapeutic antibodies are very slowly eliminated (allowing infrequent dosing, e.g., monthly injections), probably by lysosomal uptake and degradation.

Clinical Implications of Altered Bioavailability Some drugs undergo near-complete presystemic metabolism and thus cannot be administered orally. Nitroglycerin cannot be used orally because it is completely extracted prior to reaching the systemic circulation. The drug is, therefore, used by the sublingual, transdermal, or intravascular routes, which bypass presystemic metabolism.

Some drugs with very extensive presystemic metabolism can still be administered by the oral route, using much higher doses than those required intravenously. Thus, a typical intravenous dose of verapamil is 1–5 mg, compared to a usual single oral dose of 40–120 mg. Administration

of low-dose aspirin can result in exposure of cyclooxygenase in platelets in the portal vein to the drug, but systemic sparing because of first-pass aspirin deacetylation in the liver. This is an example of presystemic metabolism being exploited to therapeutic advantage.

■ PLASMA HALF-LIFE

Most pharmacokinetic processes, such as elimination, are first-order; that is, the rate of the process depends on the amount of drug present. Elimination can occasionally be zero-order (fixed amount eliminated per unit time), and this can be clinically important (see "Principles of Dose Selection," later in this chapter). In the simplest pharmacokinetic model (Fig. 67-2A), a drug bolus (D) is administered instantaneously to a central compartment, from which drug elimination occurs as a first-order process. Occasionally, central and other compartments correspond to physiologic spaces (e.g., plasma volume), whereas in other cases, they are simply mathematical functions used to describe drug disposition. The first-order nature of drug elimination leads directly to the relationship describing drug concentration (C) at any time (t) following the bolus:

$$C = \frac{D}{V_c} \cdot e^{(-0.69t/t_{1/2})}$$

where V_c is the volume of the compartment into which drug is delivered and t_{1/2} is elimination half-life. As a consequence of this relationship, a plot of the logarithm of concentration versus time is a straight line (Fig. 67-2A, inset). Half-life is the time required for 50% of a first-order process to be completed. Thus, 50% of drug elimination is achieved after one drug-elimination half-life, 75% after two, 87.5% after three, etc. In practice, first-order processes such as elimination are near-complete after four to five half-lives.

In some cases, drug is removed from the central compartment not only by elimination but also by distribution into peripheral compartments. In this case, the plot of plasma concentration versus time after a bolus may demonstrate two (or more) exponential components (Fig. 67-2B). In general, the initial rapid drop in drug concentration represents not elimination but drug distribution into and out of peripheral tissues (also first-order processes), while the slower component represents drug elimination; the initial precipitous decline is usually evident with administration by intravenous but not by other routes. Drug concentrations at peripheral sites are determined by a balance between drug distribution to and redistribution from those sites, as well as by elimination. Once distribution is near-complete (four to five distribution half-lives), plasma and tissue concentrations decline in parallel.

Clinical Implications of Half-Life Measurements The elimination half-life not only determines the time required for drug concentrations to fall to near-immeasurable levels after a single bolus, it is also the sole determinant of the time required for steady-state plasma concentrations to be achieved after any change in drug dosing (Fig. 67-4). This applies to the initiation of chronic drug therapy (whether by multiple oral doses or by continuous intravenous infusion), a change in chronic drug dose or dosing interval, or discontinuation of drug.

Steady state describes the situation during chronic drug administration when the amount of drug administered per unit time equals drug eliminated per unit time. With a continuous intravenous infusion, plasma concentrations at steady state are stable, while with chronic oral drug administration, plasma concentrations vary during the dosing interval, but the time-concentration profile between dosing intervals is stable (Fig. 67-4).

■ DRUG DISTRIBUTION

In a typical 70-kg human, plasma volume is ~3 L, blood volume is ~5.5 L, and extracellular water outside the vasculature is ~20 L. The volume of distribution of drugs extensively bound to plasma proteins but not to tissue components approaches plasma volume; warfarin is an example. By contrast, for drugs highly bound to tissues, the volume of distribution can be far greater than any physiologic space. For example, the volume of distribution of digoxin and tricyclic antidepressants is hundreds

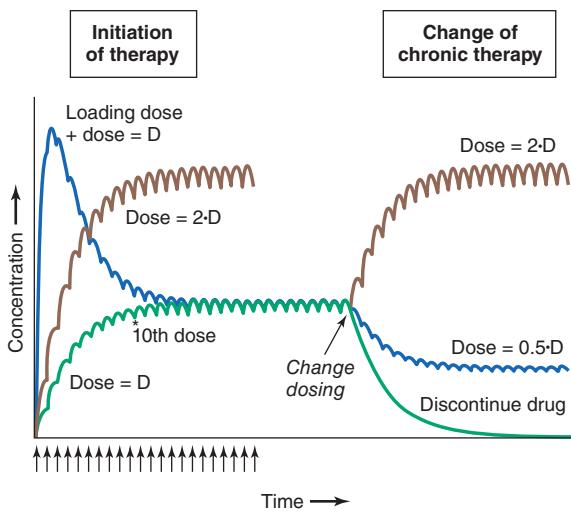


FIGURE 67-4 Drug accumulation to steady state. In this simulation, drug was administered (arrows) at intervals = 50% of the elimination half-life. Steady state is achieved during initiation of therapy after ~5 elimination half-lives, or 10 doses. A loading dose did not alter the eventual steady state achieved. A doubling of the dose resulted in a doubling of the steady state but the same time course of accumulation. Once steady state is achieved, a change in dose (increase, decrease, or drug discontinuation) results in a new steady state in ~5 elimination half-lives. (Adapted by permission from DM Roden, in DP Zipes, J Jalife [eds]: *Cardiac Electrophysiology: From Cell to Bedside*, 4th ed. Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.)

of liters, obviously exceeding total-body volume. Such drugs are not readily removed by dialysis, an important consideration in overdose.

Clinical Implications of Drug Distribution In some cases, pharmacologic effects require drug distribution to peripheral sites. In this instance, the time course of drug delivery to and removal from these sites determines the time course of drug effects; anesthetic uptake into the central nervous system (CNS) is an example.

LOADING DOSES For some drugs, the indication may be so urgent that administration of “loading” dosages is required to achieve rapid elevations of drug concentration and therapeutic effects earlier than with chronic maintenance therapy (Fig. 67-4). Nevertheless, the time required for a true steady state to be achieved is still determined only by the elimination half-life.

RATE OF INTRAVENOUS DRUG ADMINISTRATION Although the simulations in Fig. 67-2 use a single intravenous bolus, this is usually inappropriate in practice because side effects related to transiently very high concentrations can result. Rather, drugs are more usually administered orally or as a slower intravenous infusion. Some drugs are so predictably lethal when infused too rapidly that special precautions should be taken to prevent accidental boluses. For example, solutions of potassium for intravenous administration >20 mEq/L should be avoided in all but the most exceptional and carefully monitored circumstances. This minimizes the possibility of cardiac arrest due to accidental increases in infusion rates of more concentrated solutions.

Transiently high drug concentrations after rapid intravenous administration can occasionally be used to advantage. The use of midazolam for intravenous sedation, for example, depends upon its rapid uptake by the brain during the distribution phase to produce sedation quickly, with subsequent egress from the brain during the redistribution of the drug as equilibrium is achieved.

Similarly, adenosine must be administered as a rapid bolus in the treatment of reentrant supraventricular tachycardias (Chap. 246) to prevent elimination by very rapid ($t_{1/2}$ of seconds) uptake into erythrocytes and endothelial cells before the drug can reach its clinical site of action, the atrioventricular node.

Clinical Implications of Altered Protein Binding Many drugs circulate in the plasma partly bound to plasma proteins. Since only unbound (free) drug can distribute to sites of pharmacologic action,

drug response is related to the free rather than the total circulating plasma drug concentration. In chronic kidney or liver disease, protein binding may be decreased and thus drug actions increased. In some situations (myocardial infarction, infection, surgery), acute phase reactants transiently increase binding of some drugs and thus decrease efficacy. These changes assume the greatest clinical importance for drugs that are highly protein-bound since even a small change in protein binding can result in large changes in free drug; for example, a decrease in binding from 99 to 98% doubles the free drug concentration from 1 to 2%. For some drugs (e.g., phenytoin), monitoring free rather than total drug concentrations can be useful.

■ DRUG ELIMINATION

Drug elimination reduces the amount of drug in the body over time. An important approach to quantifying this reduction is to consider that drug concentrations at the beginning and end of a time period are unchanged, and that a specific volume of the body has been “cleared” of the drug during that time period. This defines clearance as volume/time. Clearance includes both drug metabolism and excretion.

Clinical Implications of Altered Clearance While elimination half-life determines the time required to achieve steady-state plasma concentration (C_{ss}), the magnitude of that steady state is determined by clearance (Cl) and dose alone. For a drug administered as an intravenous infusion, this relationship is:

$$C_{ss} = \text{dosing rate}/Cl \quad \text{or} \quad \text{dosing rate} = Cl \cdot C_{ss}$$

When a drug is administered orally, the average plasma concentration within a dosing interval ($C_{avg,ss}$) replaces C_{ss} , and the dosage (dose per unit time) must be increased if bioavailability (F) is <100%:

$$\text{Dose}/\text{time} = Cl \cdot C_{avg,ss}/F$$

Genetic variants, drug interactions, or diseases that reduce the activity of drug-metabolizing enzymes or excretory mechanisms lead to decreased clearance and, hence, a requirement for a downward dose adjustment to avoid toxicity. Conversely, some drug interactions and genetic variants increase the function of drug elimination pathways, and hence, increased drug dosage is necessary to maintain a therapeutic effect.

■ ACTIVE DRUG METABOLITES

Metabolites may produce effects similar to, overlapping with, or distinct from those of the parent drug. Accumulation of the major metabolite of procainamide, N-acetylprocainamide (NAPA), likely accounts for marked QT prolongation and torsades de pointes ventricular tachycardia (Chap. 252) during therapy with procainamide. Neurotoxicity during therapy with the opioid analgesic meperidine is likely due to accumulation of normeperidine, especially in renal disease.

Prodrugs are inactive compounds that require metabolism to generate active metabolites that mediate the drug effects. Examples include many angiotensin-converting enzyme (ACE) inhibitors, the angiotensin receptor blocker losartan, the antineoplastic irinotecan, the antiestrogen tamoxifen, the analgesic codeine (whose active metabolite morphine probably underlies the opioid effect during codeine administration), and the antiplatelet drug clopidogrel. Drug metabolism has also been implicated in bioactivation of procarcinogens and in the generation of reactive metabolites that mediate certain ADRs (e.g., acetaminophen hepatotoxicity, discussed below).

■ THE CONCEPT OF HIGH-RISK PHARMACOKINETICS

When plasma concentrations of active drug depend exclusively on a single metabolic pathway, any condition that inhibits that pathway (be it disease related, genetic, or due to a drug interaction) can lead to dramatic changes in drug concentrations and marked variability in drug action. Two mechanisms can generate highly variable drug concentrations and effects through such “high-risk pharmacokinetics.” First, variability in bioactivation of a prodrug can lead to striking variability in drug action; examples include decreased CYP2D6 activity, which prevents analgesia

by codeine, and decreased CYP2C19 activity, which reduces the antiplatelet effects of clopidogrel. The second setting is drug elimination that relies on a single pathway. In this case, inhibition of the elimination pathway by genetic variants or by administration of inhibiting drugs leads to marked elevation of drug concentration and, for drugs with a narrow therapeutic window, an increased likelihood of dose-related toxicity. The active S-enantiomer of the anticoagulant warfarin is eliminated by CYP2C9, and co-administration of amiodarone or phenytoin, CYP2C9 inhibitors, may therefore increase the risk of bleeding unless the dose is decreased. When drugs undergo elimination by multiple-drug metabolizing or excretory pathways, absence of one pathway (due to a genetic variant or drug interaction) is much less likely to have a large impact on drug concentrations or drug actions.

■ PRINCIPLES OF PHARMACODYNAMICS

Time Course of Drug Action Pharmacokinetic parameters, such as half-life and clearance, explain drug concentrations over time, but understanding the action of a drug over time (pharmacodynamics) often requires an understanding of its precise mechanism of action. Drugs act through interactions with drug targets, often in specific tissues, and with a cascade of downstream consequences. For drugs used in the urgent treatment of acute symptoms, little or no delay is anticipated (or desired) between the administration of the drug, the drug-target interaction, and the development of a clinical effect. Examples of such acute situations include vascular thrombosis, shock, or status epilepticus.

For many conditions, however, the indication for therapy is less urgent, and a delay in the onset of action clinically acceptable. Delay can be due to pharmacokinetic mechanisms such as slow elimination (resulting in slow accumulation to steady state), slow uptake into the target tissue, or slow accumulation of active metabolites. A common pharmacodynamic explanation for such a delay is the biological mechanism of action. For example, the glucocorticoid prednisolone has a plasma half-life of about 60 min. The mechanism of action, however, involves binding of the glucocorticoid receptor, translocation to the cell nucleus, and alterations in gene transcription. These downstream effects alter immune function for a much longer time frame, as evidenced by the biological half-life of 24–36 h. Other examples include proton pump inhibitors, which irreversibly bind the hydrogen/potassium adenosine triphosphatase enzyme and thus affect acid secretion for the lifetime of that enzyme, and the irreversible antiplatelet drugs, which exert effects for the duration of the life of the platelet.

Drug Effects May Be Disease Specific A drug may produce no action or a different spectrum of actions in unaffected individuals compared to patients with underlying disease. Further, concomitant disease can complicate interpretation of response to drug therapy, especially ADRs. For example, high doses of anticonvulsants such as phenytoin may cause neurologic symptoms, which may be confused with the underlying neurologic disease. Similarly, increasing dyspnea in a patient with chronic lung disease receiving amiodarone therapy could be due to the drug, underlying disease, or an intercurrent cardiopulmonary problem. As a result, alternate antiarrhythmic therapies may be preferable in patients with chronic lung disease.

While drugs interact with specific molecular receptors, drug effects may vary over time, even if stable drug and metabolite concentrations are maintained. The drug-receptor interaction occurs in a complex biologic milieu that can vary to modulate the drug effect. For example, ion channel blockade by drugs, an important anticonvulsant and antiarrhythmic effect, is often modulated by membrane potential, itself a function of factors such as extracellular potassium or local ischemia. Receptors may be up- or downregulated by disease or by the drug itself. For example, β -adrenergic blockers upregulate β -receptor density during chronic therapy. While this effect does not usually result in resistance to the therapeutic effect of the drugs, it may produce severe agonist-mediated effects (such as hypertension or tachycardia) if the blocking drug is abruptly withdrawn.

As molecular mechanisms of disease become better defined, drugs targeting those mechanisms are being introduced into practice.

Antineoplastic agents targeting mutant kinases overexpressed in cancers (e.g., BRAF V600E in melanoma, hairy cell leukemia, and other malignancies) are revolutionizing cancer care. Ivacaftor was originally developed and marketed for patients with cystic fibrosis (CF) carrying the G551D mutation in the disease gene CFTR (**Chap. 291**). While the most common CFTR mutations causing CF generate normal chloride channels that are not correctly trafficked to the cell surface, G551D channels are trafficked normally but do not conduct chloride correctly, and ivacaftor corrects this “gating” defect. Following initial marketing for only G551D patients (5% of all CF patients), the U.S. Food and Drug Administration (FDA) approved ivacaftor for use in patients carrying other CFTR mutations that confer gating defects corrected by ivacaftor *in vitro*.

■ PRINCIPLES OF DOSE SELECTION

The desired goal of therapy with any drug is to maximize the likelihood of a beneficial effect while minimizing the risk of ADRs. Previous experience with the drug, in controlled clinical trials or in postmarketing use, defines the relationships between dose or plasma concentration and these dual effects (Fig. 67-1) and has important implications for initiation of drug therapy:

1. *The target drug effect should be defined when drug treatment is started.* With some drugs, the desired effect may be difficult to measure objectively, or the onset of efficacy can be delayed for weeks or months; drugs used in the treatment of cancer and psychiatric disease are examples. Sometimes a drug is used to treat a symptom, such as pain or palpitations, and here it is the patient who will report whether the selected dose is effective. In yet other settings, such as anticoagulation or hypertension, the desired response can be repeatedly and objectively assessed by simple clinical or laboratory tests.
2. *The nature of anticipated toxicity often dictates the starting dose.* If side effects are minor, it may be acceptable to start chronic therapy at a dose highly likely to achieve efficacy and down-titrate if side effects occur. However, this approach is rarely, if ever, justified if the anticipated toxicity is serious or life-threatening; in this circumstance, it is more appropriate to initiate therapy with the lowest dose that may produce a desired effect. In cancer chemotherapy, it is common practice to use maximally tolerated doses.
3. *The above considerations do not apply if these relationships between dose and effects cannot be defined.* This is especially relevant to some ADRs (discussed further below) whose development is not readily related to drug dose.
4. *If a drug dose does not achieve its desired effect, a dosage increase is justified only if toxicity is absent and the likelihood of serious toxicity is small.*

Failure of Efficacy Even assuming the diagnosis is correct and the correct drug and dose are prescribed, drugs may fail to be effective because 100% efficacy is not expected. A complete therapeutic response is often absent with antihypertensive or antidepressant drugs, and a major challenge in contemporary therapeutics is to identify patient-specific predictors of response to individual drugs. Other explanations for failure of efficacy include drug interactions, noncompliance, or unexpectedly low drug concentration due to administration of expired or degraded drug. These are situations in which measurement of plasma drug concentrations, if available, can be especially useful. Noncompliance is an especially frequent problem in the long-term treatment of diseases such as hypertension and epilepsy, occurring in $\geq 25\%$ of patients in therapeutic environments in which no special effort is made to involve patients in the responsibility for their own health. Multidrug regimens with multiple doses per day are especially prone to noncompliance.

Monitoring response to therapy, by physiologic measures or by plasma concentration measurements, requires an understanding of the relationships between plasma concentration and anticipated effects. For example, measurement of QT interval is used during treatment with sotalol or dofetilide to avoid marked QT prolongation that can herald serious arrhythmias. In this setting, evaluating the

electrocardiogram at the time of anticipated peak plasma concentration and effect (e.g., 1–2 h postdose at steady state) is most appropriate. Maintained high vancomycin levels carry a risk of nephrotoxicity, so dosages should be adjusted on the basis of plasma concentrations measured at trough (predose). Similarly, for dose adjustment of other drugs (e.g., anticonvulsants), concentration should be measured at its lowest during the dosing interval, just prior to a dose at steady state (Fig. 67-4), to ensure a maintained therapeutic effect.

Concentration of Drugs in Plasma as a Guide to Therapy

Factors such as interactions with other drugs, disease-induced alterations in elimination and distribution, and genetic variation in drug disposition combine to yield a wide range of plasma levels in patients given the same dose. Hence, if a predictable relationship can be established between plasma drug concentration and beneficial or adverse drug effect, measurement of plasma levels can provide a valuable tool to guide selection of an optimal dose, especially when there is a narrow range between the plasma levels yielding therapeutic and adverse effects. Such therapeutic drug monitoring is commonly used with certain types of drugs including many anticonvulsants, anti-rejection agents, antiarrhythmics, and antibiotics. By contrast, if no such relationship can be established (e.g., if drug access to important sites of action outside plasma is highly variable), monitoring plasma concentration may not provide an accurate guide to therapy (Fig. 67-5).

The common situation of first-order elimination implies that average, maximum, and minimum steady-state concentrations are related linearly to the dosing rate. Accordingly, the maintenance dose may be adjusted on the basis of the ratio between the desired and measured concentrations *at steady state*; for example, if a doubling of the steady-state plasma concentration is desired, the dose should be doubled. This does not apply to drugs eliminated by zero-order kinetics (fixed amount per unit time), where small dosage increases will produce disproportionate increases in plasma concentration; examples include phenytoin and theophylline.

If an increase in dosage is needed, this is usually best achieved by increasing the drug dose and leaving the dosing interval constant

(e.g., by giving 200 mg every 8 h instead of 100 mg every 8 h). However, this approach is acceptable only if the resulting maximum concentration is not toxic and the trough value does not fall below the minimum effective concentration for an undesirable period of time. Alternatively, the steady state may be changed by altering the frequency of intermittent dosing but not the size of each dose. In this case, the magnitude of the fluctuations around the average steady-state level will change—the shorter the dosing interval, the smaller the difference between peak and trough levels.

EFFECTS OF DISEASE ON DRUG CONCENTRATION AND RESPONSE

RENAL DISEASE

Renal excretion of parent drug and metabolites is generally accomplished by glomerular filtration and by specific drug transporters. If a drug or its metabolites are primarily excreted through the kidneys and increased drug levels are associated with ADRs (an example of “high-risk pharmacokinetics” described above), drug dosages must be reduced in patients with renal dysfunction to avoid toxicity. The antiarrhythmics dofetilide and sotalol undergo predominant renal excretion and carry a risk of QT prolongation and arrhythmias if doses are not reduced in renal disease. In end-stage renal disease, sotalol has been given as 40 mg after dialysis (every second day), compared to the usual daily dose, 80–120 mg every 12 h. At approved doses, the anticoagulant edoxaban appears to be somewhat more effective in subjects with mild renal dysfunction, possibly reflecting higher drug levels. The narcotic analgesic meperidine undergoes extensive hepatic metabolism, so that renal failure has little effect on its plasma concentration. However, its metabolite, normeperidine, does undergo renal excretion, accumulates in renal failure, and probably accounts for the signs of CNS excitation, such as irritability, twitching, and seizures, that appear when multiple doses of meperidine are administered to patients with renal disease. Protein binding of some drugs (e.g., phenytoin) may be altered in uremia, so measuring free drug concentration may be desirable.

In non-end-stage renal disease, changes in renal drug clearance are generally proportional to those in creatinine clearance, which may be measured directly or estimated from the serum creatinine. This estimate, coupled with the knowledge of how much drug is normally excreted renally versus nonrenally, allows an estimate of the dose adjustment required. In practice, most decisions involving dosing adjustment in patients with renal failure use published recommended adjustments in dosage or dosing interval based on the severity of renal dysfunction indicated by creatinine clearance. Any such modification of dose is a first approximation and should be followed by plasma concentration data (if available) and clinical observation to further optimize therapy for the individual patient.

LIVER DISEASE

Standard tests of liver function are not useful in adjusting doses in diseases like hepatitis or cirrhosis. First-pass metabolism may decrease, leading to increased oral bioavailability as a consequence of disrupted hepatocyte function, altered liver architecture, and portacaval shunts. The oral bioavailability for high first-pass drugs such as morphine, meperidine, midazolam, and nifedipine is almost doubled in patients with cirrhosis, compared to those with normal liver function. Therefore, the size of the oral dose of such drugs should be reduced in this setting.

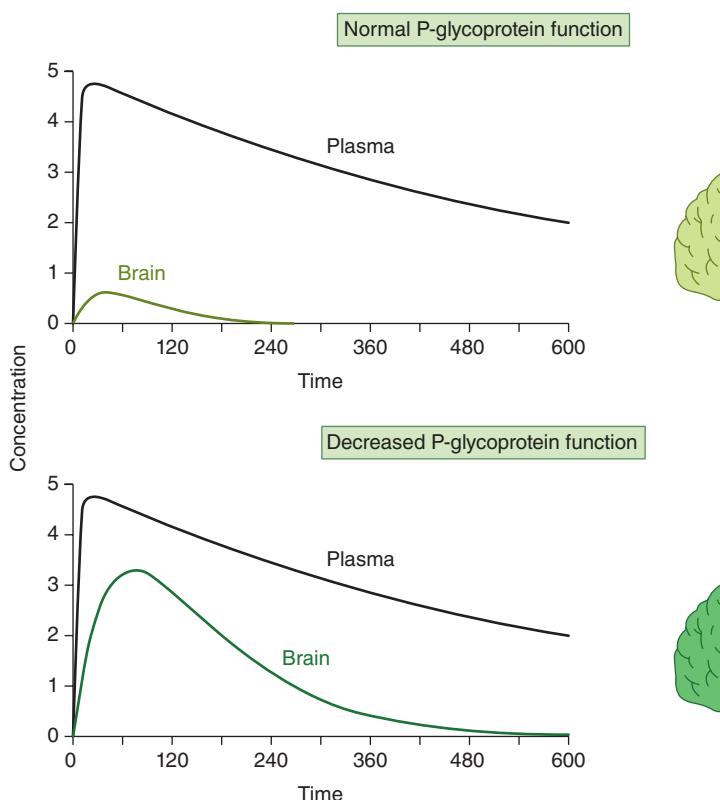


FIGURE 67-5 Drug concentrations in specific tissues may not always parallel those in plasma. For example, the efflux pump P-glycoprotein excludes drugs from the endothelium of capillaries in the brain and so constitutes a key element of the blood-brain barrier. Reduced P-glycoprotein function (e.g., due to drug interactions) can thus increase penetration of substrate drugs into the brain, even when plasma concentrations are unchanged.

HEART FAILURE AND SHOCK

Under conditions of decreased tissue perfusion, the cardiac output is redistributed to preserve blood flow to the heart and brain at the expense of other tissues (Chap. 257). As a result, drugs may be distributed into a smaller volume of distribution, higher drug concentrations will be present in the plasma, and the tissues that are best perfused (the brain and heart) will be exposed to these higher concentrations, resulting in increased CNS or cardiac effects. In addition, decreased perfusion of the kidney and liver may impair drug clearance. Another consequence of severe heart failure is decreased gut perfusion, which may reduce drug absorption and thus lead to reduced or absent effects of orally administered therapies.

DRUG USE IN THE ELDERLY

In the elderly, multiple pathologies and medications used to treat them result in more drug interactions and ADRs. Aging also results in changes in organ function, especially of the organs involved in drug disposition. Initial doses should be less than the usual adult dosage and should be increased slowly. The number of medications, and doses per day, should be kept as low as possible.

Even in the absence of kidney disease, renal clearance may be reduced by 35–50% in elderly patients. Dosages should be adjusted on the basis of creatinine clearance. Aging also results in a decrease in the size of, and blood flow to, the liver and possibly in the activity of hepatic drug-metabolizing enzymes; accordingly, the hepatic clearance of some drugs is impaired in the elderly. As with liver disease (above), these changes are not readily predicted.

Elderly patients may display altered drug sensitivity. Examples include increased analgesic effects of opioids, increased sedation from benzodiazepines and other CNS depressants, and increased risk of bleeding while receiving anticoagulant therapy, even when clotting parameters are well controlled. Exaggerated responses to cardiovascular drugs are also common because of the impaired responsiveness of normal homeostatic mechanisms. Conversely, the elderly display decreased sensitivity to β -adrenergic receptor blockers.

ADRs are especially common in the elderly because of altered pharmacokinetics and pharmacodynamics, the frequent use of multidrug regimens, and concomitant disease. For example, use of long half-life benzodiazepines is linked to the occurrence of hip fractures in elderly patients, perhaps reflecting both a risk of falls from these drugs (due to increased sedation) and the increased incidence of osteoporosis in elderly patients. In population surveys of the noninstitutionalized elderly, as many as 10% had at least one ADR in the previous year.

DRUG USE IN CHILDREN

Although there are very few pediatric-specific drugs, there are many pediatric-specific drug indications (e.g., intravenous immunoglobulin and aspirin for Kawasaki disease) and ADRs (e.g., pyloric stenosis after erythromycin exposure in infants). Drug metabolism and drug response pathways mature at different rates after birth, and the relative size of various body compartments and function of various organs change during development. There is increased motivation to avoid organ toxicity, given the anticipated long post-drug-exposure life expectancy. There are few studies providing empiric evidence to guide pediatric dosing. In practice, doses are adjusted for size (weight or body surface area) as a first approximation unless age-specific data are available. As in adults, the lowest doses anticipated to achieve clinical benefit are generally prescribed, potentially followed by titration.

INTERACTIONS BETWEEN DRUGS

Drug interactions can complicate therapy by increasing or decreasing the action of a drug; interactions may be based on changes in drug disposition or in drug response in the absence of changes in drug levels (Table 67-2). *Interactions must be considered in the differential diagnosis of any unusual response occurring during drug therapy.* Prescribers should recognize that patients often come to them with a legacy of drugs acquired during previous medical experiences, often with multiple physicians who may not be aware of all the patient's medications. A meticulous drug history should list all medications, including agents

TABLE 67-2 Drug Interactions

MECHANISM	EXAMPLE
Pharmacokinetic Interactions Causing Decreased Drug Effect	
Decreased absorption due to drug binding in the gut	Antacids or bile acid sequestrants decrease the absorption of many drugs: Antacids/tetracyclines Cholestyramine/digoxin
Decreased solubility due to altered gastric pH	H ₂ receptor blockers or proton pump inhibitors decrease solubility and absorption of weak bases: Omeprazole/ketoconazole
Induction of drug metabolism and/or drug transport:	Decreased concentrations and effects of: Warfarin Quinidine Cyclosporine Losartan Oral contraceptives Methadone Dabigatran
Decreased prodrug bioactivation	Proton pump inhibitors may prevent clopidogrel bioactivation CYP2D6 inhibitors (fluoxetine, paroxetine, quinidine, and others) may prevent codeine bioactivation
Reduced delivery of drug to active sites of action	Tricyclics prevent clonidine uptake into adrenergic neurons, preventing antihypertensive effects
Pharmacokinetic Interactions Causing Increased Drug Effect	
Inhibited drug metabolism	Cimetidine (inhibits many CYPs): Warfarin Theophylline Phenytoin CYP2D6 inhibitors ^a /β blockers CYP3A inhibitors ^a : HMG-CoA reductase inhibitors Colchicine (toxicity risk) Decreased cyclosporine dose requirement
Inhibited drug transport	Amiodarone (inhibits many CYPs and P-glycoprotein): Warfarin Digoxin Dabigatran
Inhibition of drug metabolism causing accumulation of toxic metabolites	Allopurinol (xanthine oxidase inhibitor) inhibits an alternate pathway for azathioprine and 6-mercaptopurine elimination, increasing risk for toxicity
Decreased elimination due to altered renal function	Inhibitors of renal tubular transport (phenylbutazone, probenecid, salicylates) increase methotrexate toxicity
Pharmacodynamic Drug Interactions	
Combined effects on the same biologic process	Excess bleeding with combinations of antiplatelet drugs, anticoagulants, and NSAIDs Long QT-related arrhythmias with QT-prolonging antiarrhythmics plus diuretics Hyperkalemia with ACE inhibitors plus potassium Hypotension with nitrates plus sildenafil
Antagonistic effects on the same biologic process	Loss of antihypertensive drug effects with NSAIDs

^aSee Table 67-1.

Abbreviations: ACE, angiotensin-converting enzyme; CYP, cytochrome P; NSAID, nonsteroidal anti-inflammatory drug.

not often volunteered during questioning, such as OTC drugs, health food supplements, and topical agents such as eye drops. Lists of interactions are available from a number of electronic sources. While it is unrealistic to expect the practicing physician to memorize these, certain drugs consistently run the risk of generating interactions, often by inhibiting or inducing specific drug elimination pathways; these include CYP2D6, CYP3A, and P-glycoprotein inhibitors (Table 67-1) and CYP3A/P-glycoprotein inducers (Table 67-2). Accordingly, when these drugs are started or stopped, prescribers must be especially alert to the possibility of interactions.

ADVERSE DRUG REACTIONS

The beneficial effects of drugs are coupled with the inescapable risk of untoward effects. The morbidity and mortality from these ADRs often present diagnostic problems because they can involve every organ and system of the body and may be mistaken for signs of underlying disease. In addition, some surveys have suggested that drug therapy for a range of chronic conditions such as psychiatric disease or hypertension does not achieve its desired goal in up to half of treated patients; thus, the most common “adverse” drug effect may be failure of efficacy.

ADRs can be classified in two broad groups. Type A reactions result from exaggeration of an intended pharmacologic action of the drug, such as increased bleeding with anticoagulants or bone marrow suppression with some antineoplastics, and tend to be dose-dependent. Type B reactions result from toxic effects unrelated to the intended pharmacologic actions. The latter effects are often unanticipated (especially with new drugs) and frequently severe and may result from recognized (often immunologic) as well as previously undescribed mechanisms. Type B reactions may occur at low dosages and are often termed dose-independent.

Drugs may increase the frequency of an event that is common in a general population, and this may be especially difficult to recognize; an example is the increase in myocardial infarctions that was seen with the COX-2 inhibitor rofecoxib. Drugs can also cause rare and serious ADRs, such as hematologic abnormalities, arrhythmias, severe skin reactions, or hepatic or renal dysfunction. Prior to regulatory approval and marketing, new drugs are tested in relatively few patients who tend to be less sick and to have fewer concomitant diseases than those patients who subsequently receive the drug therapeutically. Because of the relatively small number of patients studied in clinical trials and the selected nature of these patients, rare ADRs are generally not detected prior to a drug's approval; indeed, if they are detected, the new drugs are generally not approved. Therefore, physicians need to be cautious in the prescription of new drugs and alert for the appearance of previously unrecognized ADRs.

Elucidating mechanisms underlying ADRs can assist development of safer compounds or allow a patient subset at especially high risk to be excluded from drug exposure. National adverse reaction reporting systems, such as those operated by the FDA (suspected ADRs can be reported online at <http://www.fda.gov/safety/medwatch/default.htm>) and the Committee on Safety of Medicines in Great Britain, can prove useful. The publication or reporting of a newly recognized ADR can in a short time stimulate many similar such reports of reactions that previously had gone unrecognized.

Occasionally, “adverse” effects may be exploited to develop an entirely new indication for a drug. Unwanted hair growth during minoxidil treatment of severely hypertensive patients led to development of the drug for hair growth. Sildenafil was initially developed as an antianginal, but its effects to alleviate erectile dysfunction not only led to a new drug indication but also to increased understanding of the role of type 5 phosphodiesterase in erectile tissue. These examples further reinforce the concept that prescribers must remain vigilant to the possibility that unusual symptoms may reflect unappreciated drug effects.

Some 25–50% of patients make errors in self-administration of prescribed medicines, and these errors can be responsible for ADRs. Similarly, patients commit errors in taking OTC drugs by not reading or following prescribing directions on the containers. Health care providers must recognize that providing directions with prescriptions does not always guarantee compliance.

In hospitals, drugs are administered in a controlled setting, and patient compliance is, in general, ensured. Errors may occur nevertheless—the wrong drug or dose may be given or the drug may be given to the wrong patient—and improved drug distribution and administration systems should help with this problem.

SCOPE OF THE PROBLEM

One estimate in the United Kingdom was that 6.5% of all hospital admissions are due to ADRs and that 2.3% of these patients (0.15%) died as a result. The most common culprit drugs were aspirin, nonsteroidal anti-inflammatory drugs, diuretics, warfarin, ACE inhibitors, antidepressants, opiates, digoxin, steroids, and clopidogrel. One study in the late 1990s suggested that ADRs were responsible for >100,000 in-hospital deaths in the United States, making them the fourth to sixth most common cause of in-hospital death. Another study 10 years later showed no change in this trend.

In hospital, patients receive, on average, 10 different drugs during each hospitalization. The sicker the patient, the more drugs are given, and there is a corresponding increase in the likelihood of ADRs. When <6 different drugs are given to hospitalized patients, the probability of an ADR is ~5%, but if >15 drugs are given, the probability is >40%. Serious ADRs are also well recognized with “herbal” remedies and OTC compounds; examples include kava-associated hepatotoxicity, L-tryptophan-associated eosinophilia-myalgia, and phenylpropanolamine-associated stroke, each of which has caused fatalities.

TOXICITY UNRELATED TO A DRUG'S PRIMARY PHARMACOLOGIC ACTIVITY

Drugs or, more commonly, reactive metabolites generated by CYPs can covalently bind to tissue macromolecules (such as proteins or DNA) to cause tissue toxicity. Because of the reactive nature of these metabolites, covalent binding often occurs close to the site of production, typically the liver.

Acetaminophen The most common cause of drug-induced hepatotoxicity is acetaminophen overdosage (Chap. 340). Normally, reactive metabolites are detoxified by combining with hepatic glutathione. When glutathione becomes depleted, the metabolites bind instead to hepatic protein, with resultant hepatocyte damage. The hepatic necrosis produced by the ingestion of acetaminophen can be prevented or attenuated by the administration of substances such as *N*-acetylcysteine that reduce the binding of electrophilic metabolites to hepatic proteins. The risk of acetaminophen-related hepatic necrosis is increased in patients receiving drugs such as phenobarbital or phenytoin, which increase the rate of drug metabolism, or ethanol, which exhausts glutathione stores. Such toxicity has even occurred with therapeutic dosages, so patients at risk through these mechanisms should be warned.

Immunologic Reactions Most pharmacologic agents are haptens, small molecules with low molecular weights (<2000) that are therefore poor immunogens. Generation of an immune response to a drug therefore often requires *in vivo* activation and covalent linkage to protein, carbohydrate, or nucleic acid.

Drug stimulation of antibody production may mediate tissue injury by several mechanisms. The antibody may attack the drug when the drug is covalently attached to a cell and thereby destroy the cell. This occurs in penicillin-induced hemolytic anemia. Antibody-drug antigen complexes may be passively adsorbed by a bystander cell, which is then destroyed by activation of complement; this occurs in quinine- and quinidine-induced thrombocytopenia. Heparin-induced thrombocytopenia arises when antibodies against complexes of platelet factor 4 peptide and heparin generate immune complexes that activate platelets; thus, the thrombocytopenia is accompanied by “paradoxical” thrombosis and is treated with thrombin inhibitors. Drugs or their reactive metabolites may alter a host tissue, rendering it antigenic and eliciting autoantibodies. For example, hydralazine and procainamide (or their reactive metabolites) can chemically alter nuclear material, stimulating the formation of antinuclear antibodies and occasionally causing lupus erythematosus. Drug-induced pure red cell aplasia (Chap. 102) is due to an immune-based drug reaction.

Serum sickness ([Chap. 352](#)) results from the deposition of circulating drug-antibody complexes on endothelial surfaces. Complement activation occurs, chemotactic factors are generated locally, and an inflammatory response develops at the site of complex entrapment. Arthralgias, urticaria, lymphadenopathy, glomerulonephritis, or cerebritis may result. Foreign proteins (vaccines, streptokinase, therapeutic antibodies) and antibiotics are common causes. Many drugs, particularly antimicrobial agents, ACE inhibitors, and aspirin, can elicit anaphylaxis with production of IgE, which binds to mast cell membranes. Contact with a drug antigen initiates a series of biochemical events in the mast cell and results in the release of mediators that can produce the characteristic urticaria, wheezing, flushing, rhinorrhea, and (occasionally) hypotension.

Drugs may also elicit cell-mediated immune responses. One serious reaction is Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN), which can result in death due to T-cell-mediated massive skin sloughing. Another probable immune-mediated drug reaction is the DRESS (drug reaction with eosinophilia and systemic symptoms) syndrome, a rare ADR with a chronic relapsing course, often triggered by antiseizure medications and possibly arising from herpes virus reactivation. As described in [Chap. 68](#), specific genetic variants appear necessary but not sufficient to elicit SJS/TEN or DRESS.

While the use of antibodies targeting immune checkpoints is dramatically improving prognosis in many cancers, these agents have also been associated with the unpredictable development of many apparently immune-related ADRs. Some, like colitis or thyroiditis, may be self-limited or medically manageable, while others, notably myocarditis, are rarer but can be rapidly fatal.

■ DIAGNOSIS AND TREATMENT OF ADVERSE DRUG REACTIONS

The manifestations of drug-induced diseases frequently resemble those of other diseases, and a given set of manifestations may be produced by different and dissimilar drugs. Recognition of the role of a drug or drugs in an illness depends on appreciation of the possible ADRs to drugs in any disease, on identification of the temporal relationship between drug administration and development of the illness, and on familiarity with the common manifestations of the drugs.

A suspected ADR developing after introduction of a new drug naturally implicates that drug; however, it is also important to remember that a drug interaction may be responsible. Thus, for example, a patient on a chronic stable warfarin dose may develop a bleeding complication after introduction of amiodarone; this does not reflect a direct reaction to amiodarone but rather its effect to inhibit warfarin metabolism. Many associations between particular drugs and specific reactions have been described, but there is always a “first time” for a novel association, and any drug should be suspected of causing an ADR if the clinical setting is appropriate.

Illness related to a drug's intended pharmacologic action is often more easily recognized than illness attributable to immune or other mechanisms.

For example, side effects such as cardiac arrhythmias in patients receiving digitalis, hypoglycemia in patients given insulin, or bleeding in patients receiving anticoagulants are more readily related to a specific drug than are symptoms such as rash, which may be caused by many drugs or by other factors. Drug fever often escapes initial diagnosis because fever is such a common manifestation of disease.

Electronic listings of ADRs can be useful. However, exhaustive compilations often provide little sense of perspective in terms of frequency and seriousness, which can vary considerably among patients.

Eliciting a drug history from each patient is important for diagnosis. Attention must be directed to OTC drugs and herbal preparations as well as to prescription drugs. Each type can be responsible for ADRs, and adverse interactions may occur between OTC drugs and prescribed drugs. Loss of efficacy of oral contraceptives or cyclosporine with concurrent use of St. John's wort (a P-glycoprotein inducer) is an example ([Table 67-2](#)). In addition, it is common for patients to be cared for by several physicians, and duplicative, additive, antagonistic, or synergistic drug combinations may therefore be administered if

the physicians are not aware of the patients' drug histories. Electronic health records (EHRs) may help mitigate this problem, but only if all treating physicians use the same EHR system. Medications stopped for inefficacy or adverse effects should be documented to avoid pointless and potentially dangerous reexposure. A frequently overlooked source of additional drug exposure is topical therapy; for example, a patient complaining of bronchospasm may not mention that an ophthalmic beta blocker is being used unless specifically asked. A history of previous ADRs in patients is common. Since these patients have shown a predisposition to drug-induced illnesses, such a history should dictate added caution in prescribing new drugs.

Laboratory studies may include demonstration of serum antibody in some persons with drug allergies involving cellular blood elements, as in agranulocytosis, hemolytic anemia, and thrombocytopenia. For example, both quinine and quinidine can produce platelet agglutination *in vitro* in the presence of complement and the serum from a patient who has developed thrombocytopenia following use of this drug. Biochemical abnormalities such as G6PD deficiency, serum pseudocholinesterase level, or genotyping may also be useful in diagnosis, especially after an ADR has occurred in the patient or a family member ([Chap. 68](#)).

Once an ADR is suspected, discontinuation of the suspected drug followed by disappearance of the reaction is presumptive evidence of a drug-induced illness. Confirming evidence may be sought by cautiously reintroducing the drug and seeing if the reaction reappears. However, that should be done only if confirmation would be useful in the future management of the patient. Because rechallenge does carry risks, it is generally avoided unless the suspected culprit drug is critical to the patient's care. When the reaction is thought to be immunologic, challenge is generally avoided. With concentration-dependent ADRs, lowering the dosage may cause the reaction to disappear, and raising it may cause the reaction to reappear. Serious immunologically mediated ADRs have been treated with high-dose steroids; other immunosuppressive agents such as rituximab, infliximab, or mycophenolate mofetil; or plasmapheresis.

If the patient is receiving many drugs when an ADR is suspected, the drugs likeliest to be responsible can usually be identified; this should include both potential culprit agents as well as drugs that alter their elimination. All drugs may be discontinued at once or, if this is not practical, discontinued one at a time, starting with the ones most suspect, and the patient observed for signs of improvement. The time needed for a concentration-dependent ADR to disappear depends on the time required for the concentration to fall below the range associated with the ADR; that, in turn, depends on the initial blood level and on the rate of elimination or metabolism of the drug. Adverse effects of drugs with long half-lives or those not directly related to serum concentration may take a considerable time to disappear.

THE DRUG DEVELOPMENT PROCESS

Drug therapy is an ancient feature of human culture. The first treatments were plant extracts discovered empirically to be effective for indications like fever, pain, or breathlessness. This symptom-based empiric approach to drug development was supplanted in the twentieth century by identification of compounds targeting more fundamental biologic processes, such as bacterial growth or elevated blood pressure. The term “magic bullet,” coined by Paul Ehrlich to describe the search for effective compounds for syphilis, captures the essence of the hope that understanding basic biologic processes will lead to highly effective new therapies.

A common starting point for the development of many widely used modern therapies has been basic biologic discovery that implicates potential target molecules: examples of such target molecules include HMG-CoA reductase, a key step in cholesterol biosynthesis, or the *BRAF* V600E mutation that appears to drive the development of some malignant melanomas and other tumors. The development of compounds targeting these molecules has not only revolutionized treatment for diseases such as hypercholesterolemia or malignant melanoma, but has also revealed new biologic features of disease. Thus, for example, initial spectacular successes with vemurafenib (which targets

BRAF V600E were followed by near-universal tumor relapse, strongly suggesting that inhibition of this pathway alone would be insufficient for tumor control. This reasoning, in turn, supports a view that many complex diseases will not lend themselves to cure by targeting a single magic bullet, but rather single drugs or combinations that attack multiple pathways whose perturbation results in disease. The use of combination therapy in settings such as hypertension, tuberculosis, HIV infection, and many cancers highlights the potential for such a “systems biology” view of drug therapy.

A common approach in contemporary drug development is to start with a high-throughput screening procedure to identify “lead” chemical(s) modulating the activity of a potential drug target. The next step is application of increasingly sophisticated medicinal chemistry-based modification of the “lead” to develop compounds with specificity for the chosen target, lack of “off-target” effects, and pharmacokinetic properties suitable for human use (e.g., consistent bioavailability, long elimination half-life, and no high-risk pharmacokinetic features). Drug evaluation in human subjects then proceeds from initial safety and tolerance (phase 1) to dose finding (phase 2) and then to large efficacy trials (phase 3). This is a very expensive process, and the vast majority of lead compounds fail at some point. Thus, new approaches to identify likely successes and failures early are needed. One idea, described further in [Chap. 68](#), is to use genomic and other high-throughput profiling approaches not only to identify new drug targets but also to identify disease subsets for which drugs approved for other indications might be “repurposed,” thereby avoiding the costly development process.

SUMMARY

Modern clinical pharmacology aims to replace empiricism in the use of drugs with therapy based on in-depth understanding of factors that determine an individual’s response to drug treatment. Molecular pharmacology, pharmacokinetics, genetics, clinical trials, and the educated prescriber all contribute to this process. No drug response should ever be termed *idiosyncratic*; all responses have a mechanism whose understanding will help guide further therapy with that drug or successors. This rapidly expanding understanding of variability in drug actions makes the process of prescribing drugs increasingly daunting for the practitioner. However, fundamental principles should guide this process:

- The benefits of drug therapy, however defined, should always outweigh the risk.
- The smallest dosage necessary to produce the desired effect should be used.
- The number of medications and doses per day should be minimized.
- Although the literature is rapidly expanding, accessing it is becoming easier; electronic tools to search databases of literature and unbiased opinion will become increasingly commonplace.
- Genetics play a role in determining variability in drug response and may become a part of clinical practice.
- EHR and pharmacy systems will increasingly incorporate prescribing advice, such as indicated medications not used; unindicated medications being prescribed; and potential dosing errors, drug interactions, or genetically determined drug responses.
- Prescribers should be particularly wary when adding or stopping specific drugs that are especially liable to provoke interactions and adverse drug reactions.
- Prescribers should use only a limited number of drugs, with which they are thoroughly familiar.

FURTHER READING

- BARRETT JS et al: Challenges and opportunities in the development of medical therapies for pediatric populations and the role of extrapolation. *Clin Pharmacol Ther* 103:419, 2018.
- HOLFORD N: Pharmacodynamic principles and the time course of immediate drug effects. *Transl Clin Pharmacol* 25:157, 2017.
- MACRAE CA et al: The future of cardiovascular therapeutics. *Circulation* 133:2610, 2016.
- MUELLER KT et al: The role of clinical pharmacology across novel treatment modalities. *Clin Pharmacol Ther* 108:413, 2020.

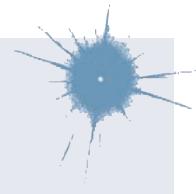
SULTANA J et al: Clinical and economic burden of adverse drug reactions. *J Pharmacol Pharmacother* 4:S73, 2013.

ZAMEK-GLISZCZYNKI MJ et al: Transporters in drug development: 2018 ITC recommendations for transporters of emerging clinical importance. *Clin Pharmacol Ther* 104:890, 2018.

68

Pharmacogenomics

Dan M. Roden



The previous chapter discussed mechanisms underlying variability in drug action, highlighting pharmacokinetic and pharmacodynamic pathways to beneficial and adverse drug events. Work in the past several decades has defined how genetic variation can play a prominent role in modulating these pathways. Initial studies described unusual drug responses due to single genetic variants in individual subjects, defining the field of pharmacogenetics. A more recent view extends this idea to multiple genetic variants across populations, and the term “pharmacogenomics” is often used. Understanding the role of genetic variation in drug response could improve the use of current drugs, avoid drug use in those at increased risk for adverse drug reactions (ADRs), guide development of new drugs, and even be used as a lens through which to understand mechanisms of diseases themselves. This chapter will outline the principles of pharmacogenomics, the evidence as currently available that genetic factors play a role in variable drug actions, and areas of controversy and ongoing work.

PRINCIPLES OF GENETIC VARIATION AND DRUG RESPONSE (SEE ALSO CHAPS. 466 AND 467)

A goal of traditional Mendelian genetics is to identify DNA variants associated with a distinct phenotype in multiple related family members ([Chap. 467](#)). However, it is unusual for a drug response phenotype to be accurately measured in more than one family member, let alone across a kindred. Some clinical studies have examined drug disposition traits (such as urinary drug excretion after a fixed test dose) in twins and have, in some instances, shown greater concordance in monozygotic compared to dizygotic pairs, supporting a genetic contribution to the trait under study. However, in general, non-family-based approaches are usually used to identify and validate DNA variants contributing to variable drug actions. Both candidate gene and genome-wide studies have been used, and as with any genomic study, results require replication before they should be accepted as valid.

Types of Genetic Variants Influencing Drug Response (Table 68-1) The most common type of genetic variant is a single nucleotide polymorphism (SNP), and nonsynonymous SNPs (i.e., those that alter primary amino acid sequence encoded by a gene) are a common cause of variant function in genes regulating drug responses, often termed *pharmacogenes*. Small insertions and deletions can similarly alter protein function or lead to functionally important splice variation. Examples of synonymous coding region variants altering pharmacogene function have also been described; the postulated mechanism is an alteration in the rate of RNA translation, and hence in folding of the nascent protein. Variation in pharmacogene promoters has been described, and copy number variation (gene deletion or multiple copies of the same gene) is also well described.

Table 68-1 lists examples of individual types of genomic variation and the impact they can have on function of pharmacogenes. Multiple genotyping approaches may be needed to detect important variants; for example, SNP assays may fail to detect large gene duplications, and highly polymorphic regions (such as the major histocompatibility locus on chromosome 6 that includes multiple genes of the human leukocyte antigen [HLA] family) are currently best evaluated by sequencing.

TABLE 68-1 Examples of Genetic Variation and Ancestry

STRUCTURAL VARIANT	EXAMPLE		FUNCTIONAL EFFECT	MINOR ALLELE FREQUENCY (%) ^a		
	COMMON NAME	dbSNP		EUROPEAN	AFRICAN	EAST ASIAN
Single nucleotide polymorphism (SNP) (or single nucleotide variant, SNV)	CYP2C9*2	rs1799853	R144C: Reduction of function	12.7	2.4	b
	CYP2C9*3	rs1057910	I359L: Loss of function	6.9	1.3	3.4
	CYP2C9*8	rs7900194	R150H: Reduction of function	b	5.6	b
	CYP2C19*2	rs4244285	Splicing defect: Loss of function	14.8	18.1	31.0
	CYP2C19*3	rs4986893	Premature stop: Loss of function	b	b	6.7
	CYP2C19*17	rs12248560	Gain of function	45	45	<5
	CYP2D6*4 ^c	rs3892097	Splicing defect: Loss of function	23.1	11.9	0.4
	CYP2D6*10 ^c		Multiple SNPs define CYP2D6*10 (reduction of function allele):			
		rs1065852	P34S	24.9	15.1	59.1
		rs1135840	S486T			
	CYP3A5*3	rs776746	Splicing defect: Loss of function	90	33	85
Insertion/deletion	VKORC1*2	rs9923231	Promoter variant associated with decreased warfarin dose	39	11	91
	VKORC1	rs61742245	D36Y: Reduction of function, associated with increased warfarin dose	5% in East Africa, Middle East, Oceania; rare elsewhere		
	ABCB1	rs1045642	Synonymous variant; may affect mRNA stability and protein folding	47.2	79.8	62.5
Insertion/deletion	UGT1A1*28		Reduction of function promoter variant (7 TA repeats versus 6 repeats in reference allele); homozygotes have Gilbert's syndrome	31.6	39.1	14.8
Multiple variants constituting specific haplotypes	HLA-B*15:01		Predispose to immunologically mediated adverse drug reactions	b	b	5
	HLA-B*57:01			6.8	1.0	1.6
Gene deletion	CYP2D6*5		Loss of function	2.7	6	5.6
Gene duplication	CYP2D6*1xN	Duplication of normal allele	Ultra-rapid metabolizer phenotype	0.8	1.5	0.3
	CYP2D6*4xN	Duplication of loss of function allele		Up to 3% in North Africa and the Middle East		
			Extensive or poor metabolizer phenotype, depending on the opposite allele	0.3	1.4	b

Note: Allele frequencies from <https://gnomad.broadinstitute.org/> and <https://cpicpgx.org/>.

^aIncludes heterozygotes and homozygotes. ^bAllele frequency <0.05%. ^cCYP2D6 is highly polymorphic, and multiple SNPs may be required to define a specific variant. For example, rs1065852 is present in both *4 and *10 variants. See <https://www.pharmvar.org/>.

Table 68-1 also highlights the fact that the frequency of important variation across pharmacogenes can vary strikingly by ancestry, with the result that certain ethnic groups may be at unusually high risk of displaying variant response to specific drugs.

Candidate Gene Approaches Most studies to date have used an understanding of the molecular mechanisms modulating drug action to identify candidate genes in which variants could explain variable drug responses. One very common scenario is that variable drug actions can be attributed to variability in plasma drug concentrations. When plasma drug concentrations vary widely (e.g., more than an order of magnitude), especially if their distribution is non-unimodal as in Fig. 68-1, variants in single genes controlling drug concentrations often contribute. In this case, the most obvious candidate genes are those responsible for drug metabolism and elimination. Other candidate genes are those encoding the target molecules with which drugs interact to produce their effects or molecules modulating that response, including those involved in disease pathogenesis.

Genome-Wide Association Studies The field has also had some success with “unbiased” approaches such as genome-wide association (GWA) (Chap. 466), particularly in identifying single variants associated with high risk for certain forms of drug toxicity, and in validating the results of candidate gene studies. GWA studies have identified variants in the HLA locus that are associated with high risk for severe skin rashes during treatment with the anticonvulsant carbamazepine and hepatotoxicity with flucloxacillin, an antibiotic never marketed in the United States. A GWA study of simvastatin-associated myopathy

identified a single noncoding SNP in *SLCO1B1*, encoding OATP1B1, a drug transporter known to modulate simvastatin uptake into the liver, which accounts for 60% of myopathy risk. African-American subjects are known to have higher dose requirements to achieve stable anticoagulation with warfarin, due in part to variations in *CYP2C9* and *VKORC1*, discussed below. In addition, a GWA study identified novel SNPs near *CYP2C9* that contribute to this effect in African Americans.

■ GENETIC VARIANTS AFFECTING PHARMACOKINETICS

Clinically important genetic variants have been described in multiple molecular pathways of drug disposition (Table 68-2). A distinct multimodal distribution of drug disposition (as shown in Fig. 68-1) argues for a predominant effect of variants in a single gene in the metabolism of that substrate. Individuals with two alleles (variants) encoding for nonfunctional protein make up one group, often termed *poor metabolizers* (PM phenotype). For most genes, many variants can produce such a loss of function, and assessing whether they are on the same or different alleles (i.e., the *diplootype*) can complicate the use of genotyping in clinical practice. Furthermore, some variants produce only partial loss of function, and the presence of more than one variant may be required to define a specific allele. Individuals with one functional allele, or multiple reduction of function alleles, make up a second group (*intermediate metabolizers*) and may or may not be distinguishable from those with two functional alleles (normal metabolizers, sometimes termed *extensive metabolizers*, EMs). *Ultra-rapid metabolizers* (UMs) with especially high enzymatic activity (occasionally due

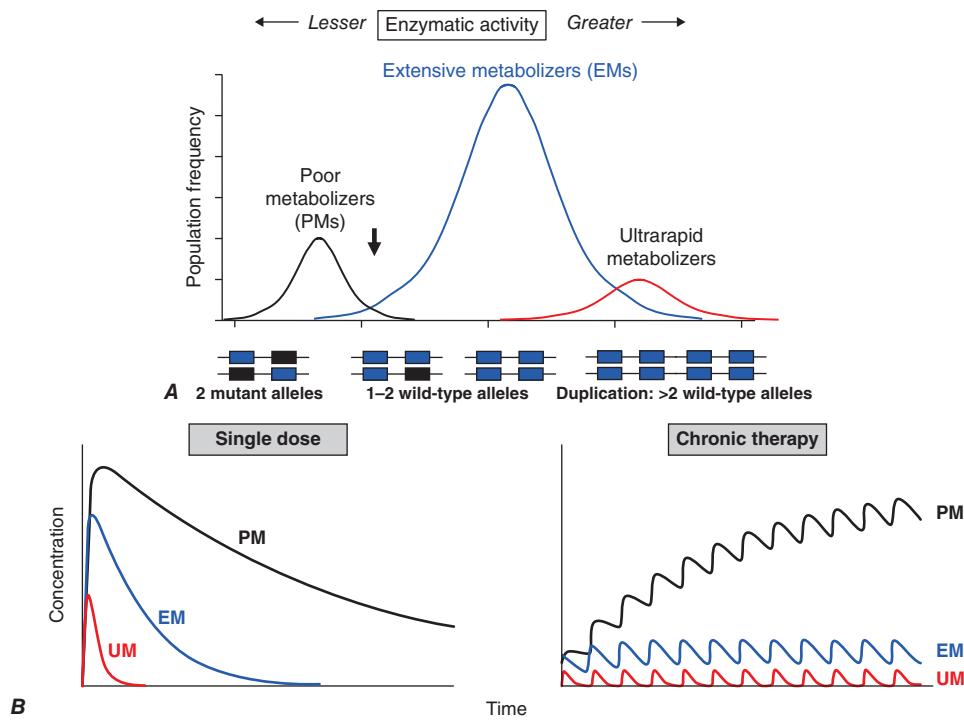


FIGURE 68-1 **A.** Distribution of CYP2D6 metabolic activity across a population. The heavy arrow indicates an antimode, separating poor metabolizer subjects (PMs, black), with two loss-of-function CYP2D6 alleles (black), indicated by the intron-exon structures below the chart. Individuals with one or two functional alleles are grouped together as extensive metabolizers (EMs, blue). Also shown are ultra-rapid metabolizers (UMs, red), with 2–12 functional copies of the gene, displaying the greatest enzyme activity. (*Adapted from M-L Dahl et al: J Pharmacol Exp Ther 274:516, 1995.*) **B.** These simulations show the predicted effects of CYP2D6 genotype on disposition of a substrate drug. With a single dose (left), there is an inverse “gene-dose” relationship between the number of active alleles and the areas under the time-concentration curves (smallest in UM subjects; highest in PM subjects); this indicates that clearance is greatest in UM subjects. In addition, elimination half-life is longest in PM subjects. The right panel shows that these single-dose differences are exaggerated during chronic therapy: steady-state concentration is much higher in PM subjects (decreased clearance), as is the time required to achieve steady state (longer elimination half-life).

TABLE 68-2 Genetic Variants and Drug Responses

GENE	DRUGS	EFFECT OF GENETIC VARIANTS ^a
Variants in Drug Metabolism Pathways		
<i>CYP2C9</i>	Losartan Warfarin Phenytoin	Decreased bioactivation and effects (PMs) Decreased dose requirements; possible increased bleeding risk (PMs) Decreased dose requirement (PMs)
<i>CYP2C19</i>	Omeprazole, voriconazole Celecoxib Clopidogrel Citalopram, escitalopram	Decreased effect in EMs Exaggerated effect in PMs Decreased effect in PMs and IMs Consider alternate drug in PMs and alternate drug or dose increase in IMs Possible increased bleeding risk in carriers of gain-of-function variants Choose alternate drug in UMs; reduce dose in PMs
<i>CYP2D6</i>	Codeine, tamoxifen Codeine Tricyclic antidepressants ^b Metoprolol, carvedilol, timolol, propafenone Fluvoxamine	Decreased bioactivation and drug effects in PMs Respiratory depression in UMs Increased adverse effects in PMs: Consider dose decrease Decreased therapeutic effects in UMs: Consider alternate drug Increased beta blockade in PMs Reduce dose or chose alternate drug in PMs
<i>CYP3A5</i>	Tacrolimus, vincristine	Decreased drug concentrations and effect (CYP3A5*3 carriers)
Dihydropyrimidine dehydrogenase (<i>DPYD</i>)	Capecitabine, 5-fluorouracil, tegafur	Possible severe toxicity (PMs)
<i>NAT2</i>	Rifampin, isoniazid, pyrazinamide, hydralazine, procainamide	Increased risk of toxicity in PMs
Thiopurine S-methyltransferase (<i>TPMT</i>)	Azathioprine, 6-mercaptopurine, thioguanine	PMs: Increased risk of bone marrow aplasia EMs: Possible decreased drug action at usual dosages
Uridine diphosphate glucuronosyltransferase (<i>UGT1A1</i>)	Irinotecan Atazanavir	PM homozygotes: Increased risk of severe adverse effects (diarrhea, bone marrow aplasia) High risk of hyperbilirubinemia during treatment; can result in drug discontinuation
Pseudocholinesterase (<i>BCHE</i>)	Succinylcholine and other muscle relaxants	Prolonged paralysis (autosomal recessive); diagnosis established by genotyping or by measuring serum cholinesterase activity

(Continued)

TABLE 68-2 Genetic Variants and Drug Responses (Continued)

GENE	DRUGS	EFFECT OF GENETIC VARIANTS*
Variants in Other Genes		
Glucose 6-phosphate dehydrogenase (<i>G6PD</i>)	Rasburicase, primaquine, chloroquine	Increased risk of hemolytic anemia in G6PD-deficient subjects
HLA-B*15:02	Carbamazepine	Carriers (1 or 2 alleles) at increased risk of SJS/TEN (mainly Asian subjects)
HLA-B*31:01	Carbamazepine	Carriers (1 or 2 alleles) at increased risk of SJS/TEN and milder skin toxicities (Caucasian and Asian subjects)
HLA-B*15:02	Phenytoin	Carriers (1 or 2 alleles) at increased risk of SJS/TEN
HLA-B*57:01	Abacavir	Carriers (1 or 2 alleles) at increased risk of SJS/TEN
HLA-B*58:01	Allopurinol	Carriers (1 or 2 alleles) at increased risk of SJS/TEN
<i>IFNL3</i> (IL28B)	Interferon	Variable response in hepatitis C therapy
<i>SLCO1B1</i>	Simvastatin	Encodes a drug uptake transporter; variant nonsynonymous single nucleotide polymorphism increases myopathy risk especially at higher dosages
<i>VKORC1</i>	Warfarin	Decreased dose requirements with variant promoter haplotype Increased dose requirement in individuals with nonsynonymous loss-of-function variants
<i>ITPA</i>	Ribavirin	Variants modulate risk for hemolytic anemia
<i>RYR1</i>	General anesthetics	Variants predispose to malignant hyperthermia
<i>CFTR</i>	Ivacaftor, lumacaftor	Targeted therapies for cystic fibrosis indicated only in certain genotypes
Variants in Other Genomes (Infectious Agents, Tumors)		
Chemokine C-C motif receptor (CCR5)	Maraviroc	Drug effective only in HIV strains with CCR5 detectable
C-KIT	Imatinib	In gastrointestinal stromal tumors, drug indicated only with c-kit-positive cases
ALK (anaplastic lymphoma kinase)	Crizotinib	Indicated in patients with non-small cell lung cancer and ALK mutations
Her2/neu overexpression	Trastuzumab, lapatinib	Drugs indicated only with tumor overexpression
K-ras mutation	Panitumumab, cetuximab	Lack of efficacy with KRAS mutation
Philadelphia chromosome	Dasatinib, nilotinib, imatinib	Decreased efficacy in Philadelphia chromosome-negative chronic myelogenous leukemia

*Drug effect in homozygotes unless otherwise specified. ^bMany tricyclic antidepressants and selective serotonin uptake inhibitors are metabolized by CYP2D6, CYP2C19, or both, and some metabolites have pharmacologic activity. See <https://www.pharmgkb.org/view/dosing-guidelines.do>.

Abbreviations: EM, extensive metabolizer (normal enzymatic activity); IM, intermediate metabolizer (heterozygote for loss-of-function allele); PM, poor metabolizer (homozygote for reduced or loss-of-function allele); SJS/TEN, Stevens-Johnson syndrome/toxic epidermal necrolysis; UM, ultra-rapid metabolizer (enzymatic activity much greater than normal, e.g., with gene duplication, Fig. 68-1).

Further data at:

U.S. Food and Drug Administration: <http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm>

Pharmacogenetics Research Network/Knowledge Base: <http://www.pharmgkb.org>

The Clinical Pharmacogenomics Implementation Consortium: <https://www.pharmgkb.org/page/cpic>

Dutch Pharmacogenetics Working Group: <https://www.knmp.nl/patientenzorg/medicatiebewaking/farmacogenetica/pharmacogenetics-1/pharmacogenetics>

to gene duplication; Table 68-1 and Fig. 68-1) have also been described for some traits. Many drugs in widespread use can inhibit specific drug disposition pathways (see Chap. 67, Table 67-1), and so EM individuals receiving such inhibitors can respond like PM patients (*phenocopying*). Polymorphisms in genes encoding drug uptake or drug efflux transporters may be other contributors to variability in drug delivery to target sites and, hence, in drug effects.

CYP3A Members of the CYP3A family (CYP3A4, CYP3A5) metabolize the greatest number of drugs in therapeutic use. CYP3A4 activity is highly variable (up to an order of magnitude) among individuals, but nonsynonymous coding region polymorphisms (those that change the encoded amino acid) are rare. Thus, the underlying mechanism likely reflects genetic variation in regulatory regions.

Most subjects of European or Asian origin carry a polymorphism that disrupts splicing in the closely related *CYP3A5* gene. As a result, these individuals display reduced CYP3A5 activity, whereas CYP3A5 activity tends to be greater in subjects of African origin. Decreased efficacy of the antirejection agent tacrolimus in subjects of African origin has been attributed to more rapid CYP3A5-mediated elimination, and a lower risk of vincristine-associated neuropathy has been reported in CYP3A5 “expressers.”

CYP2D6 CYP2D6 is second to CYP3A4 in the number of commonly used drugs that it metabolizes. CYP2D6 activity is polymorphically distributed, and 5–10% of European- and African-derived populations

(but few Asians) display the PM phenotype (Fig. 68-1). Dozens of loss-of-function variants in CYP2D6 have been described; the PM phenotype arises in individuals with two such alleles. In addition, UMs with multiple functional copies of CYP2D6 have been identified especially in East Africa, the Middle East, and Oceania. PMs have slower elimination rates and lower clearance of substrate drugs; as a consequence (Fig. 68-1B), steady-state concentrations are higher and the time taken to achieve steady state is longer than in EMs (Chap. 67). Conversely, UMs display very low steady-state parent drug concentrations and an abbreviated time to steady state.

Codeine is biotransformed by CYP2D6 to the potent active metabolite morphine, so its effects are blunted in PMs and exaggerated in UMs. Deaths due to respiratory depression in children given codeine after tonsillectomy have been attributed to the UM trait, and the U.S. Food and Drug Administration (FDA) has revised the package insert to include a prominent “black box” warning against its use in this setting, and, in fact, forbidding its use in children less than 12 years old. In the case of drugs with beta-blocking properties metabolized by CYP2D6, greater signs of beta blockade (e.g., bronchospasm, bradycardia) have been reported in PM subjects than in EMs. This can be seen not only with orally administered beta blockers such as metoprolol and carvedilol, but also with ophthalmic timolol and with the sodium channel-blocking antiarrhythmic propafenone, a CYP2D6 substrate with beta-blocking properties. UMs may require very high dosages of nortriptyline and other tricyclic antidepressants to achieve a therapeutic

effect. Tamoxifen undergoes CYP2D6-mediated biotransformation to an active metabolite, so its efficacy may be in part related to this polymorphism. In addition, the widespread use of selective serotonin reuptake inhibitors (SSRIs) to treat tamoxifen-related hot flashes may also alter the drug's effects because many SSRIs, notably fluoxetine and paroxetine, are also CYP2D6 inhibitors (Table 67-2).

CYP2C19 The PM phenotype for CYP2C19 is common (20%) among Asians and rarer (2–3%) in other populations; the frequency of the PM trait is especially high (>50%) in Oceania. The impact of polymorphic CYP2C19-mediated metabolism has been demonstrated with the proton pump inhibitor omeprazole, where ulcer cure rates with “standard” dosages were much lower in EM patients (29%) than in PMs (100%). Thus, understanding the importance of this polymorphism would have been important in developing the drug, and knowing a patient's CYP2C19 genotype could improve therapy. CYP2C19 is responsible for bioactivation of the antiplatelet drug clopidogrel, and several large retrospective, and more recently prospective, studies have documented decreased efficacy (e.g., increased myocardial infarction after placement of coronary stents or increased stroke or transient ischemic attacks) among subjects with one or two reductions of function alleles. In addition, some studies suggest that omeprazole and possibly other proton pump inhibitors phenocopy this effect by inhibiting CYP2C19.

CYP2C9 There are common variants in CYP2C9 that encode proteins with reduction or loss of catalytic function. These variant alleles are associated with increased rates of neurologic complications with phenytoin, hypoglycemia with glipizide, and reduced warfarin dose required to maintain stable anticoagulation. Rare patients homozygous for loss-of-function alleles may require very low warfarin dosages. Up to 50% of the variability in steady-state warfarin dose requirement is attributable to polymorphisms in CYP2C9 and in the promoter of VKORC1, which encodes the warfarin target with lesser contributions by genes such as CYP4F2 controlling vitamin K metabolism. The angiotensin receptor blocker losartan is a prodrug that is bioactivated by CYP2C9; as a result, PMs and those receiving inhibitor drugs may display little response to therapy.

DPYD Individuals homozygous for loss-of-function alleles in dihydropyrimidine dehydrogenase, encoded by DPYD, are at increased risk for severe toxicity when exposed to the substrate anticancer drug 5-fluorouracil (5-FU), as well as to capecitabine and tegafur, which are metabolized to 5-FU. Dose reductions have been recommended in intermediate metabolizers.

Transferase Variants Thiopurine S-methyltransferase (TPMT) bioactivates the antileukemic drug 6-mercaptopurine (6-MP), and 6-MP is itself an active metabolite of the immunosuppressive azathioprine. Homozygotes for alleles encoding inactive TPMT (1/300 individuals) predictably exhibit severe and potentially fatal pancytopenia on standard doses of azathioprine or 6-MP. On the other hand, homozygotes for fully functional alleles may display less anti-inflammatory or antileukemic effect with standard doses of the drugs. GWA studies have also identified loss-of-function variants in NUDT15 that reduce degradation of thiopurine metabolites and, thereby, also increase risk of excessive myelosuppression.

N-acetylation is accomplished by hepatic N-acetyl transferase (NAT), which represents the activity of two genes, NAT1 and NAT2. Both enzymes transfer an acetyl group from acetyl coenzyme A to the drug; polymorphisms in NAT2 are thought to underlie individual differences in the rate at which drugs are acetylated and thus define “rapid acetylators” and “slow acetylators.” Slow acetylators make up ~50% of European and African populations but are less common among East Asians. Slow acetylators have an increased incidence of the drug-induced lupus syndrome during procainamide and hydralazine therapy and of hepatitis with isoniazid.

Individuals homozygous for a common promoter polymorphism that reduces transcription of uridine diphosphate glucuronosyltransferase (UGT1A1) have benign hyperbilirubinemia (Gilbert's syndrome;

Chap. 337). This variant has also been associated with diarrhea and increased bone marrow depression with the antineoplastic prodrug irinotecan, whose active metabolite is normally detoxified by UGT1A1-mediated glucuronidation. The antiretroviral atazanavir is a UGT1A1 inhibitor, and individuals with the Gilbert's variant develop higher bilirubin levels during treatment. While this is benign, the hyperbilirubinemia can complicate clinical care because it may raise the question of whether coexistent hepatic injury is present.

Transporter Variants The risk for myotoxicity with simvastatin and possibly other statins appears increased with variants in SLCO1B1. Variants in ABCB1, encoding the drug efflux transporter P-glycoprotein, may increase digoxin toxicity. Variants in the uptake transporters MATE1 and MATE2 have been reported to modulate metformin's glucose-lowering activity.

■ GENETIC VARIANTS AFFECTING PHARMACODYNAMICS

A variant in the VKORC1 promoter, especially common in Asian subjects (Table 68-1), reduces transcriptional activity and warfarin dose requirement. Multiple polymorphisms identified in the β_2 -adrenergic receptor appear to be linked to specific drug responses in asthma and congestive heart failure, diseases in which β_2 -receptor function might be expected to determine drug response. Polymorphisms in the β_2 -receptor gene have also been associated with response to inhaled β_2 -receptor agonists, while those in the β_1 -adrenergic receptor gene have been associated with variability in heart rate slowing and blood pressure lowering. In addition, in heart failure, the arginine allele of the common β_1 -adrenergic receptor gene polymorphism R389G has been associated with decreased mortality and decreased incidence of atrial fibrillation during treatment with the investigational beta blocker bucindolol.

Drugs may also interact with genetic pathways of disease to elicit or exacerbate symptoms of the underlying conditions. In the porphyrias, CYP inducers are thought to increase the activity of enzymes proximal to the deficient enzyme, exacerbating or triggering attacks (Chap. 416). Deficiency of glucose-6-phosphate dehydrogenase (G6PD), most often in individuals of African, Mediterranean, or South Asian descent, increases the risk of hemolytic anemia in response to the antimalarial primaquine (Chap. 100) and the uric acid-lowering agent rasburicase, which does not cause hemolysis in patients with normal amounts of the enzyme. Patients with mutations in RYR1 encoding the skeletal muscle intracellular release calcium (also termed type 1 ryanodine receptor) are asymptomatic until exposed to certain general anesthetics, which can trigger the rare syndrome of malignant hyperthermia. Certain antiarrhythmics and other drugs can produce marked QT prolongation and torsades de pointes (Chap. 246), and in a minority of affected patients, this adverse effect represents unmasking of previously subclinical congenital long QT syndrome.

Immunologically Mediated Drug Reactions The Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) is a potentially fatal skin and systemic reaction now increasingly recognized to be linked to specific HLA alleles (Table 68-2). Cases of drug-induced hepatotoxicity and of the drug rash with eosinophilia and systemic symptoms (DRESS) syndrome have also been linked to variants in this region. The frequency of risk alleles often varies by ancestry (Table 68-1). The HLA risk alleles appear to be necessary but not sufficient to elicit these reactions. For example, HLA-B*57:01 is a risk allele for abacavir-related SJS/TEN and flucloxacillin-related hepatotoxicity. However, while 55% of abacavir-exposed subjects will develop a reaction, only 1/10,000 subjects exposed to flucloxacillin develop hepatotoxicity. Thus, a third factor, the nature of which has not yet been established, seems necessary.

Tumor and Infectious Agent Genomes The actions of drugs used to treat infectious or neoplastic disease may be modulated by variants in these nonhuman germline genomes. Genotyping tumors is a rapidly evolving approach to target therapies to underlying mechanisms and to avoid potentially toxic therapy in patients who would derive

no benefit (**Chap. 71**). Trastuzumab, which potentiates anthracycline-related cardiotoxicity, is ineffective in breast cancers that do not express the Herceptin receptor. Imatinib targets a specific tyrosine kinase, BCR-Abl1, that is generated by the translocation that creates the Philadelphia chromosome typical of chronic myelogenous leukemia (CML). Imatinib is also an inhibitor of another kinase, c-kit, and the drug is remarkably effective in c-kit-driven cancer, such as gastrointestinal stromal tumors (**Chap. 71**). Vemurafenib does not inhibit wild-type *BRAF* but is active against the V600E mutant form of the kinase. Crizotinib is highly effective in non-small cell lung cancers harboring anaplastic lymphoma kinase (ALK) mutations.

■ INCORPORATING PHARMACOGENETIC INFORMATION INTO CLINICAL PRACTICE

The discovery of common variant alleles with relatively large effects on drug response raises the prospect that these variants could be used to guide therapy. Desired outcomes could be better ways of choosing likely effective drugs and dosages, or avoiding drugs that are likely to produce severe adverse drug events or be ineffective in individual subjects. Indeed, the FDA now incorporates pharmacogenetic data into package inserts meant to guide prescribing. A decision to adopt pharmacogenetically guided dosing for a given drug depends on multiple factors. The most important are the magnitude and clinical importance of the genetic effect and the strength of evidence linking genetic variation to variable drug effects (e.g., anecdote versus post-hoc analysis of clinical trial data versus randomized clinical trial [RCT]). The evidence can be strengthened if statistical arguments from clinical trial data are complemented by an understanding of underlying physiologic mechanisms. Cost versus expected benefit may also be a factor.

Point of Care Versus Preemptive Approaches Two approaches to pharmacogenetic implementation have been put in place at “early adopter” institutions and are currently being evaluated. In the first, variant-specific assays are ordered at the time of drug prescription and delivered rapidly (often within an hour or two), and the results are then used to guide therapy with that specific drug. The alternative to this “point-of-care” approach is a “preemptive” approach in which pharmacogenetic testing for large numbers of potential variants across many drugs is undertaken prior to prescription of any such drug. The data are then available in electronic health record (EHR) systems and coupled to real-time clinical decision support (CDS). When a drug whose effects are known to be influenced by pharmacogenetic variants is prescribed, the EHR system looks up whether variants likely to affect response are present; if so, CDS will alert health care providers that an alternate drug or a different dose may be required.

Challenges There are multiple challenges in putting in place either system. Assay validity and reproducibility have been issues in the past, but are less likely now. National consortia are now being put in place to develop standards for pharmacogenetic CDS. While common variants in genes such as those listed in Table 68-1 have been clearly associated with variable drug responses, the effect of rare variants, now readily discoverable by large-scale sequencing, is unknown. The extent to which a dose adjustment might be recommended may vary depending on whether zero, one, or two variant alleles are present, and whether such variants are reduction of function, loss of function, or gain of function. The Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Dutch Pharmacogenetics Working Group have developed and published guidelines for multiple drug-gene pairs focusing on the question of what might be an appropriate drug dose adjustment given the availability of genetic data. These resources do not directly address the question of when or how such genetic testing should be undertaken.

Developing Evidence That Pharmacogenetic Testing Alters Drug Outcomes A major issue is whether pharmacogenetic testing affects important drug response outcomes. When the evidence is compelling, alternate therapies are not available, and there are clear recommendations for dosage adjustment in subjects with variants, there is a strong argument for deploying genetic testing as a guide to

prescribing; HLA-B*57:01 testing for abacavir is an example described below. In other situations, the arguments are less compelling: the magnitude of the genetic effect may be smaller, the consequences may be less serious, alternate therapies may be available, or the drug effect may be amenable to monitoring by other approaches.

One school argues that the physiology and pharmacology are known and that RCTs are, therefore, unnecessary (and conceivably unethical). The analogy is sometimes drawn to well-recognized dose adjustment of renally excreted drugs in the presence of renal dysfunction. RCTs have not been conducted and the idea of such dose adjustment is well accepted in the medical community and recommended in FDA-approved drug labels. Others have argued that the effect of genetic variants is generally modest and variability in drug actions has many nongenetic sources, so genetic testing might provide marginal benefit at best.

Efforts to demonstrate the value of pharmacogenetic testing have met with mixed results. An RCT clearly showed that HLA-B*57:01 testing eliminates SJS/TEN due to abacavir. Similarly, regulatory authorities in some countries in Southeast Asia mandated HLA-B*15:02 testing prior to initiation of carbamazepine; however, in this case, an unfortunate outcome in some jurisdictions was that prescribers stopped using carbamazepine, often substituting phenytoin (another drug associated with SJS/TEN), so the incidence of the severe ADR was unchanged.

RCTs evaluating the effect of using pharmacogenetically guided therapy to optimize warfarin treatment have shown either no effect or a modest benefit of incorporating genetic information into prescribing the drug. Initial RCTs focused on time in therapeutic range in the first 4–12 weeks of treatment, whereas one more recent trial demonstrated that genotype-guided therapy could reduce the frequency of over-anticoagulation. Retrospective analyses of bleeding cases versus non-bleeding controls in EHRs and administrative databases have suggested a role for CYP2C9*3 or for the V433M variant in CYP4F2 in mediating this risk.

Two large trials have randomized patients with acute coronary syndromes to newer antiplatelet therapies (ticagrelor or prasugrel) or clopidogrel if *CYP2C19* variants were absent; in one, clopidogrel was superior, and in the second, a trend in the same direction, which did not reach the prespecified endpoint, was observed.

New effective alternate therapies to warfarin and clopidogrel that appear to lack important pharmacogenetic variants have emerged. One approach to therapy, therefore, is to use pharmacogenetic testing to identify subjects in whom variants are absent and therefore standard doses of the conventional inexpensive drugs are likely to be effective and to reserve alternate more expensive therapies for subjects likely to have variant responses to warfarin or clopidogrel.

■ GENETICS AND DRUG DEVELOPMENT

Genetic tools are now being increasingly used to identify or validate new drug targets. Initial studies suggest that a new drug development program is more likely to succeed if evidence from human genetics supports the role of a possible drug target in disease pathogenesis and suggests that the risk of toxicity due to high-risk pharmacokinetics or other mechanisms is small. Furthermore, studies of the relationships between variants in genes encoding drug target molecules and a range of phenotypes (e.g., those in EHRs) are being used for drug “repurposing,” identifying new indications for existing drugs.

Finding Protective Alleles Can Identify Drug Targets One example of using genetics to identify a new drug target started with the discovery that very rare gain-of-function variants in *PCSK9* are a rare cause of familial hypercholesterolemia. Subsequently, population studies showed that carriers of loss-of-function SNPs (2.5% of African Americans) had decreased low-density lipoprotein cholesterol, decreased incidence of coronary artery disease, and no deleterious consequences in other organ systems. These data triggered the development of PCSK9 monoclonal antibodies, which were marketed <10 years after the initial population studies. Other targets implicated by similar population genetic studies include HSD17B13 for

prevention of chronic liver disease, SLC30A8 for the prevention of type 2 diabetes, and APOC3 for hypertriglyceridemia. Discovering rare protective alleles may require very large data sets (>100,000), such as EHR systems coupled to DNA biobanks or epidemiologic cohorts like the UK Biobank.

Cancer In cancer, tumor sequencing has identified new targets for drug development, often constitutively active kinases. A problem in this area has been the rapid emergence of drug resistance, often after extraordinary initial responses. For example, 40% of melanomas appear to be driven by the V600E mutant form of *BRAF*, and the specific inhibitor vemurafenib can produce clinically spectacular remission. However, durable responses are rare, and it is now apparent that combination therapy, often with inhibitors of the MEK pathway, can provide improved therapy. Another approach that is rapidly gaining wide use in cancer involves drugs that reverse immune system inhibition (Chap. 73). In some patients, the release of this “brake” can provide durable remissions, whereas in others, severe adverse events, including colitis, pneumonitis, and myocarditis, have been reported. Understanding the mechanisms underlying variability to these therapies is a major emerging challenge in the field.

Using Multiple Data Types The development of methods to understand associations across multiple large data sets is another approach that is being explored in drug development. For example, a GWA study of risk of rheumatoid arthritis identified multiple risk loci, and many encode proteins that are known targets for intervention in the disease. Interestingly, others encode proteins that are targets for drugs used in other conditions, such as certain cancers, raising the question of whether such drugs could be “repurposed” for rheumatoid arthritis.

While the field has, to date, focused on individual high effect size variants (that are often common in a population), newer approaches combining many (dozens to millions) common variants into polygenic risk scores to predict drug responses are also being explored. An extension of this approach is the broader issue of systems pharmacology, in which multiple sources of data are used to identify potential molecules or pathways that would be amenable to treatment, by new

drugs or by existing agents, using analysis of genomic, transcriptomic, proteomic, and other large data sets. Similar approaches are being developed to predict toxicity expected from targeting specific genes or disease pathways.

SUMMARY

The science of pharmacogenomics has evolved from isolated examples of rare adverse drug actions to a more comprehensive view of the role of genetic variation in mediating the effects of most drugs. Current principles include:

- Genetic variants with an important effect on drug actions can be common, and their frequencies often vary by ancestry.
- One common mechanism is modulation of drug concentrations.
- No practitioner can be expected to remember all variants important for all drugs. Electronic data systems can now be accessed to describe this information. Ultimately, this information will be used by linking individual pharmacogenetic data to smart EHR systems.
- Incorporating genetic approaches into drug development projects holds the promise of more rapid development of targeted, safe, and effective therapies.

FURTHER READING

- CHENOWETH MJ et al: Global pharmacogenomics within precision medicine: Challenges and opportunities. *Clin Pharmacol Ther* 107:57, 2020.
- DOIGO D et al: Phenome-wide association studies across large population cohorts support drug target validation. *Nat Commun* 9:4285, 2018.
- LUZUM JA et al: The Pharmacogenomics Research Network Translational Pharmacogenetics Program: Outcomes and metrics of pharmacogenetic implementations across diverse healthcare systems. *Clin Pharmacol Ther* 102:502, 2017.
- OSANLOU O et al: Pharmacogenetics of adverse drug reactions. *Adv Pharmacol* 83:155, 2018.
- RELLING MV et al: The clinical pharmacogenetics implementation consortium: 10 years later. *Clin Pharmacol Ther* 107:171, 2020.
- RODEN DM et al: Pharmacogenomics. *Lancet* 394:521, 2019.

Section 1 Neoplastic Disorders

69

Approach to the Patient with Cancer

Dan L. Longo



The application of current treatment techniques (surgery, radiation therapy, chemotherapy, and biologic therapy) results in the cure of nearly two of three patients diagnosed with cancer. Nevertheless, patients experience the diagnosis of cancer as one of the most traumatic and revolutionary events that has ever happened to them. Independent of prognosis, the diagnosis brings with it a change in a person's self-image and in his or her role in the home and workplace. The prognosis of a person who has just been found to have pancreatic cancer is the same as the prognosis of the person with aortic stenosis who develops the first symptoms of congestive heart failure (median survival, ~8 months). However, the patient with heart disease may remain functional and maintain a self-image as a fully intact person with just a malfunctioning part, a diseased organ ("a bum ticker"). By contrast, the patient with pancreatic cancer has a completely altered self-image and is viewed differently by family and anyone who knows the diagnosis. He or she is being attacked and invaded by a disease that could be anywhere in the body. Every ache or pain takes on desperate

significance. Cancer is an exception to the coordinated interaction among cells and organs. In general, the cells of a multicellular organism are programmed for collaboration. Many diseases occur because the specialized cells fail to perform their assigned task. Cancer takes this malfunction one step further. Not only is there a failure of the cancer cell to maintain its specialized function, but it also strikes out on its own; the cancer cell competes to survive using natural mutability and natural selection to seek advantage over normal cells in a recapitulation of evolution. One consequence of the traitorous behavior of cancer cells is that the patient feels betrayed by his or her body. The cancer patient feels that he or she, and not just a body part, is diseased.

THE MAGNITUDE OF THE PROBLEM

No nationwide cancer registry exists; therefore, the incidence of cancer is estimated on the basis of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database, which tabulates cancer incidence and death figures from 13 sites, accounting for about 10% of the U.S. population, and from population data from the U.S. Census Bureau. In 2021, 1,898 million new cases of invasive cancer (970,250 men and 927,910 women) were diagnosed, and 608,570 persons (319,420 men and 289,150 women) died from cancer. The percent distribution of new cancer cases and cancer deaths by site for men and women is shown in **Table 69-1**. Cancer incidence has been declining by about 2% each year since 1992. Cancer is the cause of one in four deaths in the United States.

The most significant risk factor for cancer overall is age; two-thirds of all cases were in those aged >65 years. Cancer incidence increases as the third, fourth, or fifth power of age in different sites. For the interval

TABLE 69-1 Distribution of Cancer Incidence and Deaths for 2021

MALE			FEMALE		
SITES	%	NUMBER	SITES	%	NUMBER
Cancer Incidence					
Prostate	26	248,530	Breast	30	281,550
Lung	12	119,100	Lung	13	116,660
Colorectal	8	79,520	Colorectal	8	69,980
Bladder	7	64,280	Endometrial	7	66,570
Melanoma	6	62,260	Melanoma	5	43,850
Kidney	5	48,780	Lymphoma	4	35,930
Lymphoma	5	45,630	Thyroid	3	32,130
Oral cavity	4	38,800	Pancreas	3	28,480
Leukemia	4	35,530	Kidney	3	27,300
Pancreas	3	31,950	Leukemia	3	25,560
All others	20	195,870	All others	21	199,900
All sites	100	970,250	All sites	100	927,910
Cancer Deaths					
Lung	22	69,410	Lung	22	62,470
Prostate	11	34,130	Breast	15	43,600
Colorectal	9	28,520	Colorectal	8	24,460
Pancreas	8	25,270	Pancreas	8	22,950
Liver	6	20,300	Ovary	5	14,460
Leukemia	4	13,900	Endometrial	4	12,940
Esophagus	4	12,410	Liver	3	9,930
Bladder	4	12,260	Leukemia	3	9,760
Lymphoma	4	12,170	Lymphoma	3	8,550
CNS	3	10,500	CNS	3	8,100
All others	25	80,550	All others	25	71,930
All sites	100	319,420	All sites	100	289,150

Source: From Cancer Statistics 2021, RI Siegel et al, © 2021 CA Cancer J Clin. Reproduced with permission of John Wiley & Sons Ltd.

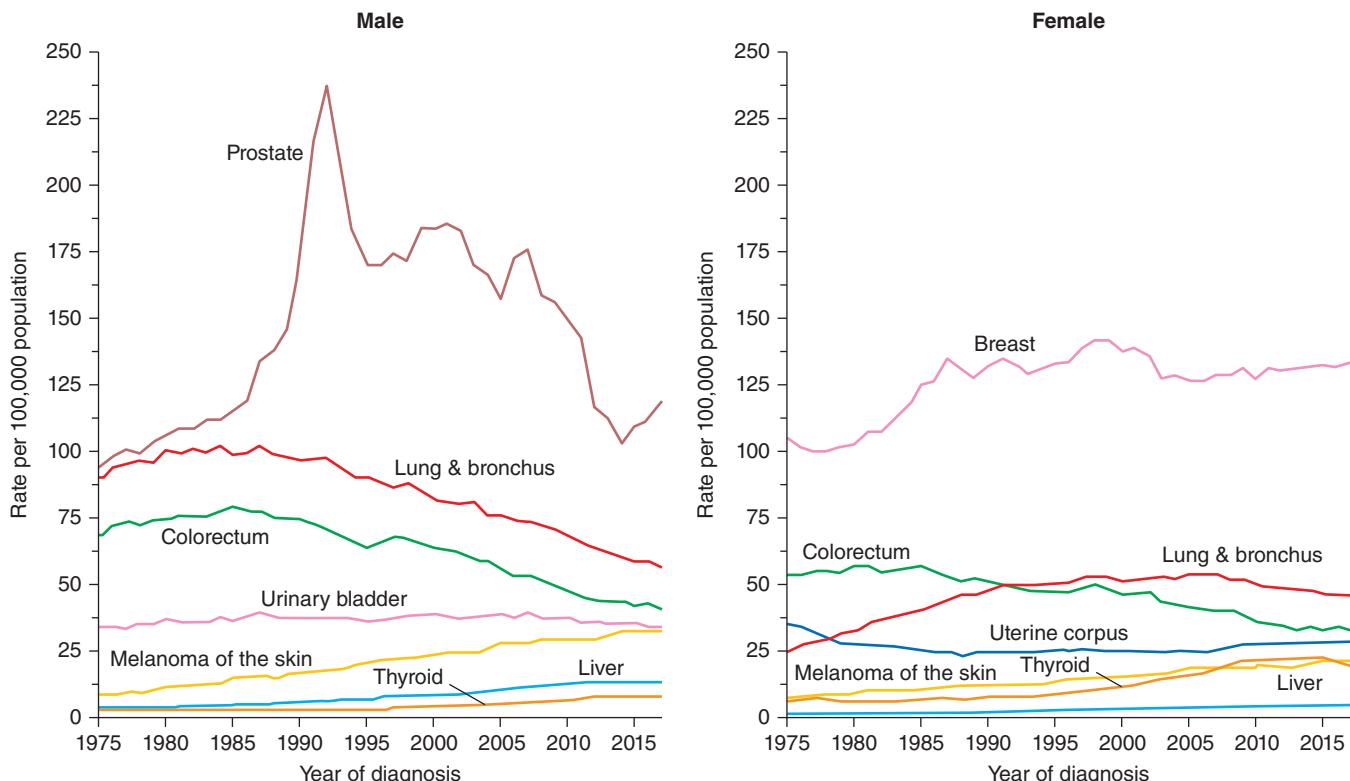


FIGURE 69-1 Trends in cancer incidence, 1975–2017. (From Cancer Statistics 2021, RI Siegel et al, © 2021 CA Cancer J Clin. Reproduced with permission of John Wiley & Sons Ltd.)

between birth and age 49 years, 1 in 29 men and 1 in 19 women will develop cancer; for the interval between ages 50 and 59 years, 1 in 15 men and 1 in 17 women will develop cancer; for the interval between ages 60 and 69 years, 1 in 6 men and 1 in 10 women will develop cancer; and for people aged ≥ 70 , 1 in 3 men and 1 in 4 women will develop cancer. Overall, men have a 40.5% risk of developing cancer at some time during their lives; women have a 38.9% lifetime risk.

Cancer is the second leading cause of death behind heart disease. Deaths from heart disease have declined 45% in the United States since 1950 and continue to decline. Cancer has overtaken heart disease as the number one cause of death in persons aged < 85 years. Incidence trends over time are shown in Fig. 69-1. After a 70-year period of increase, cancer deaths began to decline in 1990–1991 (Fig. 69-2). Between 1990 and 2010, cancer deaths decreased by 21% among men and 12.3% among women. The incidence has been steady since 2013. The magnitude of the decline is illustrated in Fig. 69-3. The five leading causes of cancer deaths are shown for various populations in Table 69-2. The 5-year survival for white patients was 39% in 1960–1963 and 68% in 2010–2016. Cancers are more often deadly in blacks; the 5-year survival was 63% for the 2010–2016 interval; however, the racial differences are narrowing over time. Incidence and mortality vary among racial and ethnic groups (Table 69-3). The basis for these differences is unclear.

Advances in cancer prevention, diagnosis, and treatment since the early 1990s have averted millions of cancer deaths based on projections from the slopes of the mortality curves leading up to the 1990s (Fig. 69-4).

CANCER AROUND THE WORLD

In 2018, 17 million new cancer cases and 9.5 million cancer deaths were estimated worldwide, according to estimates of GLOBOCAN 2018, developed by the International Agency for Research on Cancer (IARC). Rates are increasing worldwide. When broken down by region of the world, ~45% of cases were in Asia (which has 59.5% of the world's population), 26% in Europe (9.8% of the world's population), 14.5% in North America, 7.1% in Central/South America (the Americas, North and South, account for 13.3% of the world's population), 6% in Africa (16.9% of the world's population), and 1% in Australia/New Zealand

(0.5% of the world's population) (Fig. 69-5). Lung cancer is the most common cancer and the most common cause of cancer death in the world. Its incidence is highly variable, affecting only 2 per 100,000 African women but as many as 61 per 100,000 North American men. Breast cancer is the second most common cancer worldwide; however, it ranks fourth as a cause of death behind lung, stomach, and liver cancer. Among the eight most common forms of cancer, lung (2-fold), breast (3-fold), prostate (2.5-fold), and colorectal (3-fold) cancers are more common in more developed countries than in less developed countries. By contrast, liver (2-fold), cervical (2-fold), and esophageal (2- to 3-fold) cancers are more common in less developed countries. Stomach cancer incidence is similar in more and less developed countries but is much more common in Asia than North America or Africa. The most common cancers in Africa are cervical, breast, and liver cancers. It has been estimated that nine modifiable risk factors are responsible for more than one-third of cancers worldwide. These include smoking, alcohol consumption, obesity, physical inactivity, low fruit and vegetable consumption, unsafe sex, air pollution, indoor smoke from household fuels, and contaminated injections.

PATIENT MANAGEMENT

Important information is obtained from every portion of the routine history and physical examination. The duration of symptoms may reveal the chronicity of disease. The past medical history may alert the physician to the presence of underlying diseases that may affect the choice of therapy or the side effects of treatment. The social history may reveal occupational exposure to carcinogens or habits, such as smoking or alcohol consumption, that may influence the course of disease and its treatment. The family history may suggest an underlying familial cancer predisposition and point out the need to begin surveillance or other preventive therapy for unaffected siblings of the patient. The review of systems may suggest early symptoms of metastatic disease or a paraneoplastic syndrome.

DIAGNOSIS

The diagnosis of cancer relies most heavily on invasive tissue biopsy and should never be made without obtaining tissue; no noninvasive

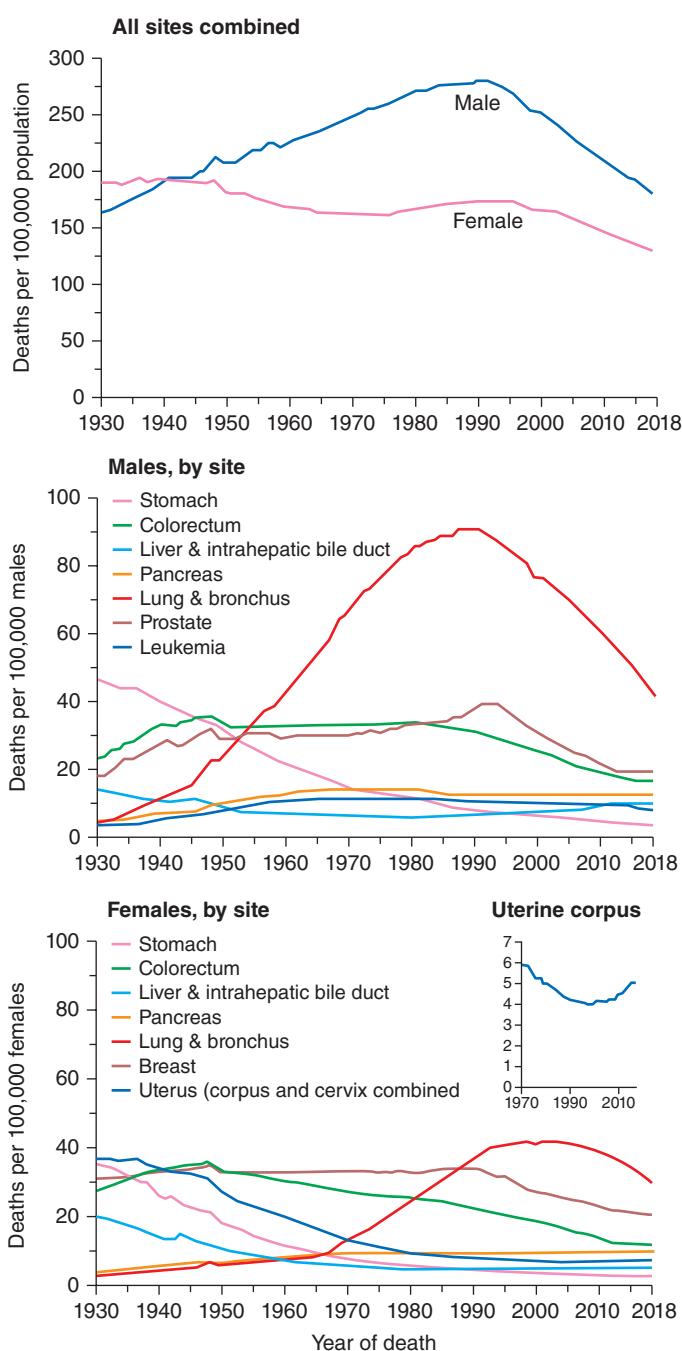


FIGURE 69-2 Trends in cancer mortality rates in men and women, 1930–2018. (From Cancer Statistics 2021, R. L. Siegel et al., © 2021 CA Cancer J Clin. Reproduced with permission of John Wiley & Sons Ltd.)

diagnostic test is sufficient to define a disease process such as cancer. Although in rare clinical settings (e.g., thyroid nodules), fine-needle aspiration is an acceptable diagnostic procedure, the diagnosis generally depends on obtaining adequate tissue to permit careful evaluation of the histology of the tumor, its grade, and its invasiveness and to yield further molecular diagnostic information, such as the expression of cell-surface markers or intracellular proteins that typify a particular cancer, or the presence of a molecular marker, such as the t(8;14) translocation of Burkitt's lymphoma. Increasing evidence links the expression of certain genes with the prognosis and response to therapy (Chaps. 71 and 72).

Occasionally, a patient will present with a metastatic disease process that is defined as cancer on biopsy but has no apparent primary site of disease. Efforts should be made to define the primary site based on age, sex, sites of involvement, histology and tumor markers, and personal and family history. Particular attention should be focused on ruling out the most treatable causes (Chap. 92).

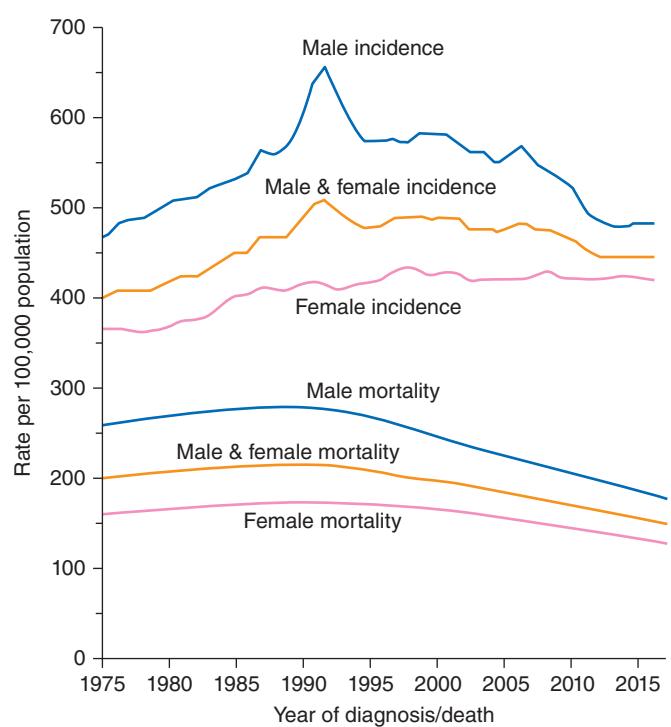


FIGURE 69-3 Trends in cancer incidence and death rates. (From Cancer Statistics 2021, R. L. Siegel et al., © 2021 CA Cancer J Clin. Reproduced with permission of John Wiley & Sons Ltd.)

Once the diagnosis of cancer is made, the management of the patient is best undertaken as a multidisciplinary collaboration among the primary care physician, medical oncologists, surgical oncologists, radiation oncologists, oncology nurse specialists, pharmacists, social workers, rehabilitation medicine specialists, and a number of other consulting professionals working closely with each other and with the patient and family.

DEFINING THE EXTENT OF DISEASE AND THE PROGNOSIS

The first priority in patient management after the diagnosis of cancer is established and shared with the patient is to determine the extent of disease. The curability of a tumor usually is inversely proportional to the tumor burden. Ideally, the tumor will be diagnosed before symptoms develop or as a consequence of screening efforts (Chap. 70). A very high proportion of such patients can be cured. However, most patients with cancer present with symptoms related to the cancer, caused either by mass effects of the tumor or by alterations associated with the production of cytokines or hormones by the tumor.

For most cancers, the extent of disease is evaluated by a variety of noninvasive and invasive diagnostic tests and procedures. This process is called *staging*. There are two types. *Clinical staging* is based on physical examination, radiographs, isotopic scans, computed tomography (CT) scans, and other imaging procedures; *pathologic staging* takes into account information obtained during a surgical procedure, which might include intraoperative palpation, resection of regional lymph nodes and/or tissue adjacent to the tumor, and inspection and biopsy of organs commonly involved in disease spread. Pathologic staging includes histologic examination of all tissues removed during the surgical procedure. Surgical procedures performed may include a simple lymph node biopsy or more extensive procedures such as thoracotomy, mediastinoscopy, or laparotomy. Surgical staging may occur in a separate procedure or may be done at the time of definitive surgical resection of the primary tumor. A subset of pathologic staging is the examination of tissue obtained at initial surgery that occurs after the delivery of some treatment, which is called neoadjuvant therapy. Stage of disease determined after neoadjuvant therapy is designated with the prefix *y*.

Knowledge of the predilection of particular tumors for spreading to adjacent or distant organs helps direct the staging evaluation.

TABLE 69-2 The Five Leading Primary Tumor Sites for Patients Dying of Cancer Based on Age and Sex in 2018

RANK	SEX	ALL AGES	AGE, YEARS				
			UNDER 20	20–39	40–59	60–79	>80
1	M	Lung	CNS	CNS	Lung	Lung	Lung
	F	Lung	CNS	Breast	Breast	Lung	Lung
2	M	Prostate	Leukemia	Colorectal	Colorectal	Prostate	Prostate
	F	Breast	Leukemia	Cervix	Lung	Breast	Breast
3	M	Colorectal	Bone sarcoma	Leukemia	Liver	Pancreas	Colorectal
	F	Colorectal	Soft tissue sarcoma	Colorectal	Colorectal	Pancreas	Colorectal
4	M	Pancreas	Soft tissue sarcoma	Lymphoma	Pancreas	Colorectal	Bladder
	F	Pancreas	Bone sarcoma	CNS	Ovary	Colorectal	Pancreas
5	M	Liver	Lymphoma	Soft tissue sarcoma	CNS	Liver	Pancreas
	F	Ovary	Kidney	Leukemia	Pancreas	Ovary	Leukemia

Abbreviations: CNS, central nervous system; F, female; M, male.

Source: From RL Siegel et al: Cancer statistics, 2021. CA Cancer J Clin 71:7, 2021.

Information obtained from staging is used to define the extent of disease as localized, as exhibiting spread outside of the organ of origin to regional but not distant sites, or as metastatic to distant sites. The most widely used system of staging is the tumor, node, metastasis (TNM) system codified by the International Union Against Cancer and the American Joint Committee on Cancer. The TNM classification is an anatomically based system that categorizes the tumor on the basis of the size of the primary tumor lesion (T1–4, where a higher number indicates a tumor of larger size), the presence of nodal involvement

(usually N0 and N1 for the absence and presence, respectively, of involved nodes, although some tumors have more elaborate systems of nodal grading), and the presence of metastatic disease (M0 and M1 for the absence and presence, respectively, of metastases). The various permutations of T, N, and M scores (sometimes including tumor histologic grade [G]) are then broken into stages, usually designated by the roman numerals I through IV. Tumor burden increases and curability decreases with increasing stage. Other anatomic staging systems are used for some tumors, e.g., the Dukes classification for

TABLE 69-3 Cancer Incidence and Mortality in Racial and Ethnic Groups, United States, 2013–2018

SITE	SEX	WHITE	BLACK	ASIAN/PACIFIC ISLANDER	AMERICAN INDIAN ^a	HISPANIC
Incidence per 100,000 Population						
All	M	501.4	534.0	294.3	399.8	371.3
	F	442.2	406.6	292.6	388.8	335.5
Breast		131.6	127.3	95.6	94.9	94.8
Colorectal	M	42.6	51.6	34.6	47.2	39.9
	F	31.8	37.9	24.8	38.3	27.6
Kidney	M	23.1	26.1	11.2	31.3	21.9
	F	11.7	13.3	5.3	17.7	12.4
Liver	M	10.7	18.0	19.3	22.9	20.1
	F	3.8	5.5	7.1	9.4	7.9
Lung	M	70.8	79.8	43.2	59.2	37.1
	F	56.4	47.9	27.9	47.9	24.3
Prostate		97.7	171.6	53.8	67.7	85.6
Cervix		7.2	9.0	6.1	8.8	9.5
Deaths per 100,000 Population						
All	M	190.2	227.2	114.6	169.3	134.0
	F	137.8	154.9	84.6	120.1	94.6
Breast		20.1	28.2	11.7	14.8	13.8
Colorectal	M	16.1	23.2	11.2	18.5	14.0
	F	11.5	15.3	7.9	12.4	8.6
Kidney	M	5.5	5.5	2.5	8.3	4.9
	F	2.3	2.3	1.1	3.2	2.2
Liver	M	8.4	13.4	13.1	14.8	13.3
	F	3.6	4.9	5.4	7.0	6.0
Lung	M	49.4	57.0	28.0	38.4	23.0
	F	35.6	30.6	16.3	27.4	12.3
Prostate		17.9	38.3	8.8	18.5	15.6
Cervix		2.0	3.4	1.7	2.4	2.6

^aBased on Indian Health Service delivery areas.

Abbreviations: F, female; M, male.

Source: From Cancer Statistics 2021, RL Siegel et al, © 2021 CA Cancer J Clin. Reproduced with permission of John Wiley & Sons Ltd.

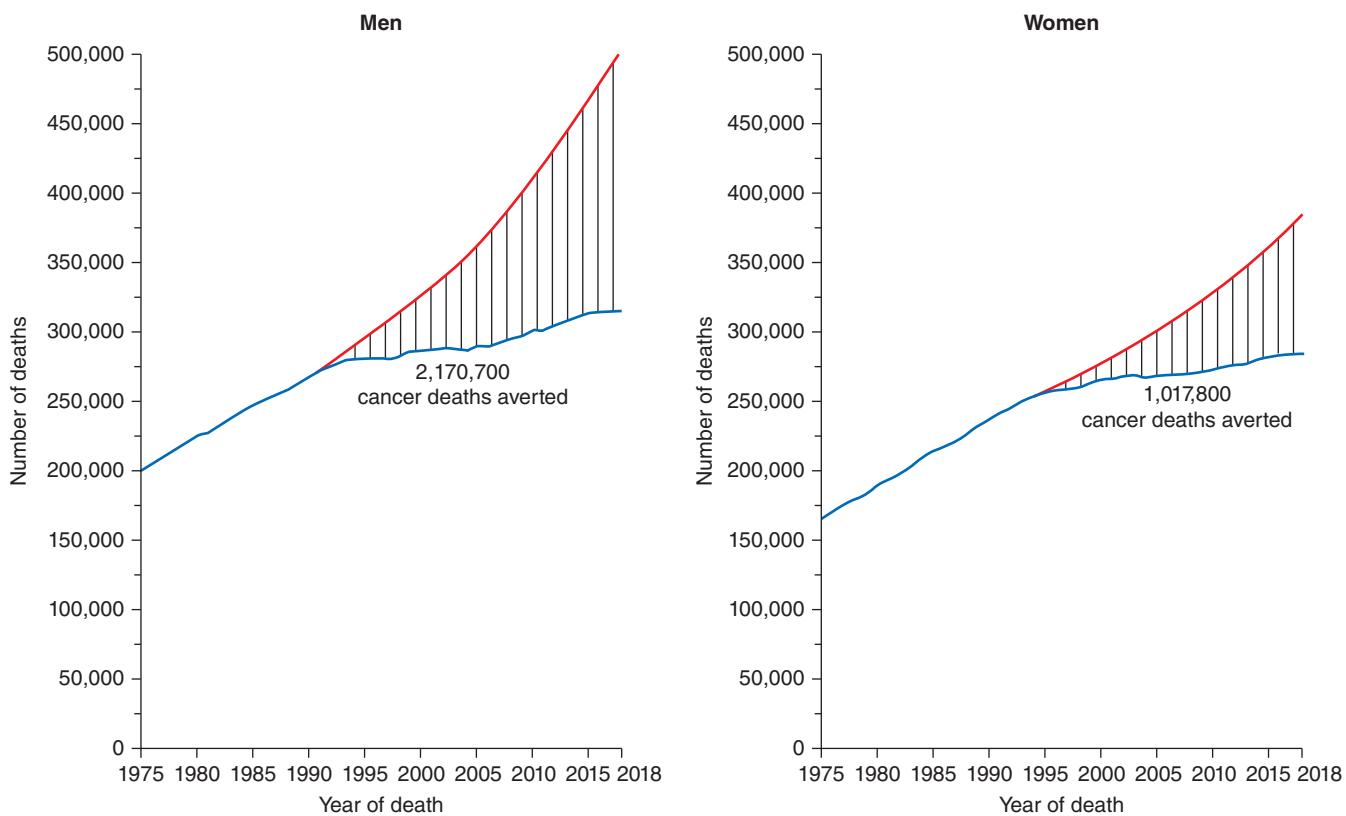


FIGURE 69-4 Cancer deaths averted in men and women since the early 1990s. (From *Cancer Statistics 2021*, RI Siegel et al, © 2021 CA Cancer J Clin. Reproduced with permission of John Wiley & Sons Ltd.)

colorectal cancers, the International Federation of Gynecologists and Obstetricians classification for gynecologic cancers, and the Ann Arbor classification for Hodgkin's disease.

Certain tumors cannot be grouped on the basis of anatomic considerations. For example, hematopoietic tumors such as leukemia, myeloma, and lymphoma are often disseminated at presentation and do not spread like solid tumors. For these tumors, other prognostic factors have been identified (*Chaps. 104–111*).

In addition to tumor burden, a second major determinant of treatment outcome is the physiologic reserve of the patient. Patients who are bedridden before developing cancer are likely to fare worse, stage for stage, than fully active patients. Physiologic reserve is a determinant of how a patient is likely to cope with the physiologic stresses imposed by the cancer and its treatment. This factor is difficult to assess directly. Instead, surrogate markers for physiologic reserve are used, such as the patient's age or Karnofsky performance status (*Table 69-4*) or Eastern Cooperative Oncology Group (ECOG) performance status (*Table 69-5*). Older patients and those with a Karnofsky performance status <70 or ECOG performance status ≥3 have a poor prognosis unless the poor performance is a reversible consequence of the tumor.

Increasingly, biologic features of the tumor are being related to prognosis. The expression of particular oncogenes, drug-resistance genes, apoptosis-related genes, and genes involved in metastasis is being found to influence response to therapy and prognosis. The presence of selected cytogenetic abnormalities may influence survival. Tumors with higher growth fractions, as assessed by expression of proliferation-related markers such as proliferating cell nuclear antigen, behave more aggressively than tumors with lower growth fractions. Information obtained from studying the tumor itself will increasingly be used to influence treatment decisions. Host genes involved in drug metabolism can influence the safety and efficacy of particular treatments.

Enormous heterogeneity has been noted by studying tumors; we have learned that morphology is not capable of discerning certain distinct subsets of patients whose tumors have different sets of abnormalities. Tumors that look the same by light microscopy can be very different. Similarly, tumors that look quite different from one another histologically can share genetic lesions that predict responses to

treatments. Furthermore, tumor cells vary enormously within a single patient even though the cells share a common origin.

MAKING A TREATMENT PLAN

From information on the extent of disease and the prognosis and in conjunction with the patient's wishes, it is determined whether the treatment approach should be curative or palliative in intent. Cooperation among the various professionals involved in cancer treatment is of the utmost importance in treatment planning. For some cancers, chemotherapy or chemotherapy plus radiation therapy delivered before the use of definitive surgical treatment (so-called neoadjuvant therapy) may improve the outcome, as seems to be the case for locally advanced breast cancer and head and neck cancers. In certain settings in which combined-modality therapy is intended, coordination among the medical oncologist, radiation oncologist, and surgeon is crucial to achieving optimal results. Sometimes the chemotherapy and radiation therapy need to be delivered sequentially, and other times concurrently. Surgical procedures may precede or follow other treatment approaches. It is best for the treatment plan either to follow a standard protocol precisely or else to be part of an ongoing clinical research protocol evaluating new treatments. Ad hoc modifications of standard protocols are likely to compromise treatment results.

The choice of treatment approaches was formerly dominated by the local culture in both the university and the practice settings. However, it is now possible to gain access electronically to standard treatment protocols and to every approved clinical research study in North America through a personal computer interface with the Internet.¹

¹The National Cancer Institute maintains a database called PDQ (Physician Data Query) that is accessible on the Internet under the name CancerNet at <https://www.cancer.gov/publications/pdq>. Information can be obtained through a facsimile machine using CancerFax by dialing 301-402-5874. Patient information is also provided by the National Cancer Institute in at least three formats: on the Internet via CancerNet at www.cancer.gov, through the CancerFax number listed above, or by calling 1-800-4-CANCER. The quality control for the information provided through these services is rigorous.

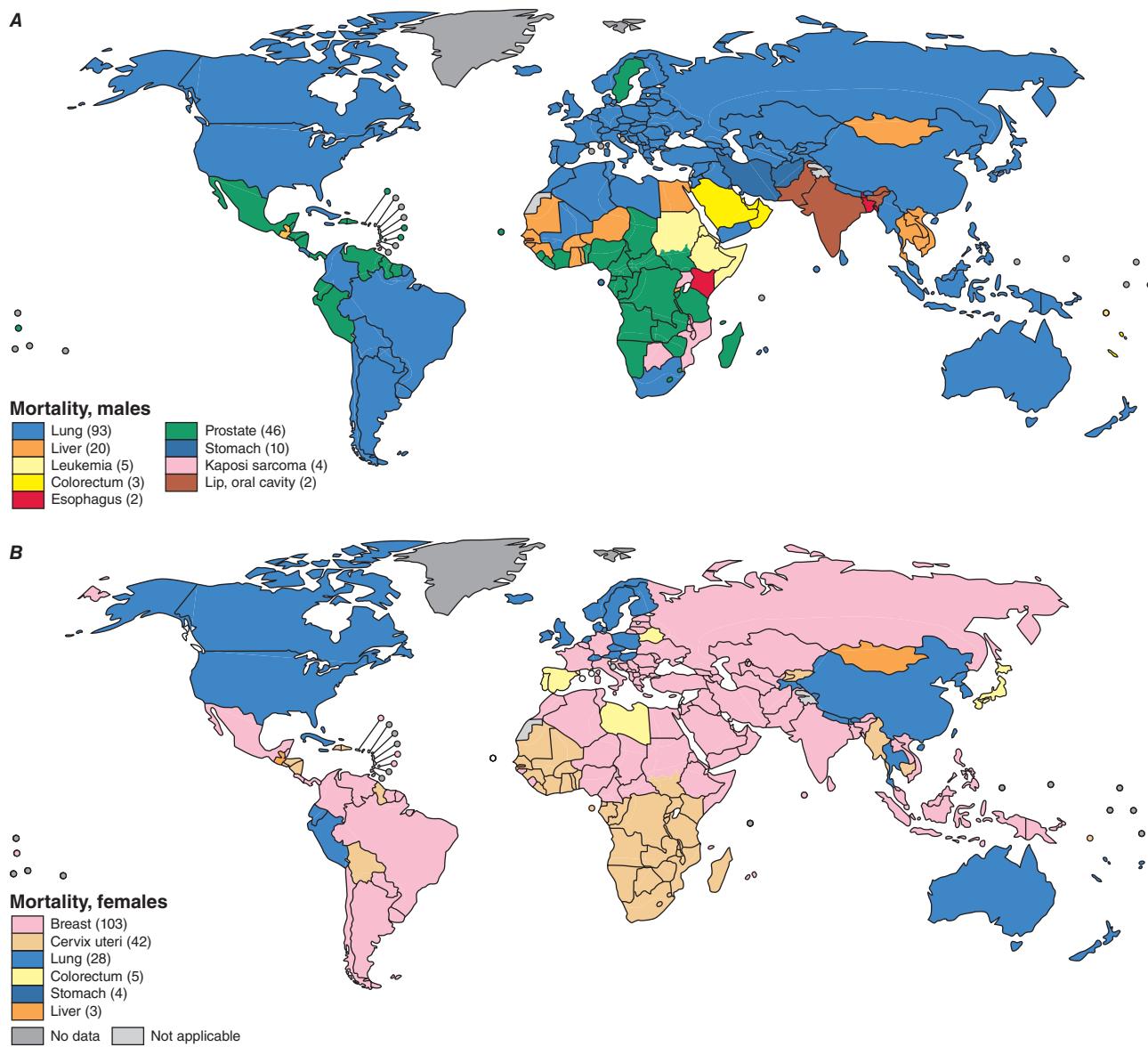


FIGURE 69-5 Global maps showing most common cause of cancer mortality by country in 2018 among (A) men and (B) women. (Reproduced with permission from F Bray et al: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394, 2018. Data source: Globocan 2018. Map production: IARC. World Health Organization. © WHO 2018. All rights reserved.)

TABLE 69-4 Karnofsky Performance Index

PERFORMANCE STATUS	FUNCTIONAL CAPABILITY OF THE PATIENT
100	Normal; no complaints; no evidence of disease
90	Able to carry on normal activity; minor signs or symptoms of disease
80	Normal activity with effort; some signs or symptoms of disease
70	Cares for self; unable to carry on normal activity or do active work
60	Requires occasional assistance but is able to care for most needs
50	Requires considerable assistance and frequent medical care
40	Disabled; requires special care and assistance
30	Severely disabled; hospitalization is indicated, although death is not imminent
20	Very sick; hospitalization is necessary; active supportive treatment is necessary
10	Moribund, fatal processes progressing rapidly
0	Dead

The skilled physician also has much to offer the patient for whom curative therapy is no longer an option. Often a combination of guilt and frustration over the inability to cure the patient and the pressure of a busy schedule greatly limit the time a physician spends with a patient who is receiving only palliative care. Resist these forces. In addition

TABLE 69-5 The Eastern Cooperative Oncology Group (ECOG) Performance Scale

ECOG grade 0: Fully active, able to carry on all predisease performance without restriction
ECOG grade 1: Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light housework, office work
ECOG grade 2: Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about >50% of waking hours
ECOG grade 3: Capable of only limited self-care, confined to bed or chair >50% of waking hours
ECOG grade 4: Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair
ECOG grade 5: Dead

Source: Reproduced with permission from MM Oken et al: Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 5:649, 1982.

to the medicines administered to alleviate symptoms (see below), it is important to remember the comfort that is provided by holding the patient's hand, continuing regular examinations, and taking time to talk.

■ MANAGEMENT OF DISEASE AND TREATMENT COMPLICATIONS

Because cancer therapies are toxic (Chap. 73), patient management involves addressing complications of both the disease and its treatment as well as the complex psychosocial problems associated with cancer. In the short term during a course of curative therapy, the patient's functional status may decline. Treatment-induced toxicity is less acceptable if the goal of therapy is palliation. The most common side effects of treatment are nausea and vomiting (see below), febrile neutropenia (Chap. 74), and myelosuppression (Chap. 73). Tools are now available to minimize the acute toxicity of cancer treatment.

New symptoms developing in the course of cancer treatment should always be assumed to be reversible until proven otherwise. The fatalistic attribution of anorexia, weight loss, and jaundice to recurrent or progressive tumor could result in a patient dying from a reversible intercurrent cholecystitis. Intestinal obstruction may be due to reversible adhesions rather than progressive tumor. Systemic infections, sometimes with unusual pathogens, may be a consequence of the immunosuppression associated with cancer therapy. Some drugs used to treat cancer or its complications (e.g., nausea) may produce central nervous system symptoms that look like metastatic disease or may mimic paraneoplastic syndromes such as the syndrome of inappropriate antidiuretic hormone. A definitive diagnosis should be pursued and may even require a repeat biopsy.

A critical component of cancer management is assessing the response to treatment. In addition to a careful physical examination in which all sites of disease are physically measured and recorded in a flow chart by date, response assessment usually requires periodic repeating of imaging tests that were abnormal at the time of staging. If imaging tests have become normal, repeat biopsy of previously involved tissue is performed to document complete response by pathologic criteria. Biopsies are not usually required if there is macroscopic residual disease. A *complete response* is defined as disappearance of all evidence of disease, and a *partial response* as >50% reduction in the sum of the products of the perpendicular diameters of all measurable lesions. The determination of partial response may also be based on a 30% decrease in the sums of the longest diameters of lesions (Response Evaluation Criteria in Solid Tumors [RECIST]). *Progressive disease* is defined as the appearance of any new lesion or an increase of >25% in the sum of the products of the perpendicular diameters of all measurable lesions (or an increase of 20% in the sums of the longest diameters by RECIST). Tumor shrinkage or growth that does not meet any of these criteria is considered *stable disease*. Some sites of involvement (e.g., bone) or patterns of involvement (e.g., lymphangitic lung or diffuse pulmonary infiltrates) are considered unmeasurable. No response is complete without biopsy documentation of their resolution, but partial responses may exclude their assessment unless clear objective progression has occurred.

For some hematologic neoplasms, flow cytometric and genetic assays may determine the presence of residual tumor cells that escape microscopic detection. In general, these techniques can reliably detect as few as 1 tumor cell among 10,000 cells. If such tests do not detect tumor cells, the patient is said to have minimal residual disease negativity, a finding generally associated with more durable remissions. Accumulating data are defining interventions in patients with minimal residual disease positivity that can extend remission duration and survival.

Tumor markers may be useful in patient management in certain tumors. Response to therapy may be difficult to gauge with certainty. However, some tumors produce or elicit the production of markers that can be measured in the serum or urine, and in a particular patient, rising and falling levels of the marker are usually associated with increasing or decreasing tumor burden, respectively. Some clinically useful tumor markers are shown in Table 69-6. Tumor markers are not

TABLE 69-6 Tumor Markers

TUMOR MARKERS	CANCER	NONNEOPLASTIC CONDITIONS
Hormones		
Human chorionic gonadotropin	Gestational trophoblastic disease, gonadal germ cell tumor	Pregnancy
Calcitonin	Medullary cancer of the thyroid	
Catecholamines	Pheochromocytoma	
Oncofetal Antigens		
α Fetoprotein	Hepatocellular carcinoma, gonadal germ cell tumor	Cirrhosis, hepatitis
Carcinoembryonic antigen	Adenocarcinomas of the colon, pancreas, lung, breast, ovary	Pancreatitis, hepatitis, inflammatory bowel disease, smoking
Enzymes		
Prostatic acid phosphatase	Prostate cancer	Prostatitis, prostatic hypertrophy
Neuron-specific enolase	Small-cell cancer of the lung, neuroblastoma	
Lactate dehydrogenase	Lymphoma, Ewing's sarcoma	Hepatitis, hemolytic anemia, many others
Tumor-Associated Proteins		
Prostate-specific antigen	Prostate cancer	Prostatitis, prostatic hypertrophy
Monoclonal immunoglobulin	Myeloma	Infection, MGUS
CA-125	Ovarian cancer, some lymphomas	Menstruation, peritonitis, pregnancy
CA 19-9	Colon, pancreatic, breast cancer	Pancreatitis, ulcerative colitis
CD30	Hodgkin's disease, anaplastic large-cell lymphoma	—
CD25	Hairy cell leukemia, adult T-cell leukemia/lymphoma	Hemophagocytic lymphohistiocytosis

Abbreviation: MGUS, monoclonal gammopathy of uncertain significance.

in themselves specific enough to permit a diagnosis of malignancy to be made, but once a malignancy has been diagnosed and shown to be associated with elevated levels of a tumor marker, the marker can be used to assess response to treatment.

The recognition and treatment of depression are important components of management. The incidence of depression in cancer patients is ~25% overall and may be greater in patients with greater debility. This diagnosis is likely in a patient with a depressed mood (dysphoria) and/or a loss of interest in pleasure (anhedonia) for at least 2 weeks. In addition, three or more of the following symptoms are usually present: appetite change, sleep problems, psychomotor retardation or agitation, fatigue, feelings of guilt or worthlessness, inability to concentrate, and suicidal ideation. Patients with these symptoms should receive therapy. Medical therapy with a serotonin reuptake inhibitor such as fluoxetine (10–20 mg/d), sertraline (50–150 mg/d), or paroxetine (10–20 mg/d) or a tricyclic antidepressant such as amitriptyline (50–100 mg/d) or desipramine (75–150 mg/d) should be tried, allowing 4–6 weeks for response. Effective therapy should be continued at least 6 months after resolution of symptoms. If therapy is unsuccessful, other classes of antidepressants may be used. In addition to medication, psychosocial interventions such as support groups, psychotherapy, and guided imagery may be of benefit.

Many patients opt for unproven or unsound approaches to treatment when it appears that conventional medicine is unlikely to be curative. Those seeking such alternatives are often well educated and

may be early in the course of their disease. Unsound approaches are usually hawked on the basis of unsubstantiated anecdotes and not only cannot help the patient but may be harmful. Physicians should strive to keep communications open and nonjudgmental, so that patients are more likely to discuss with the physician what they are actually doing. The appearance of unexpected toxicity may be an indication that a supplemental therapy is being taken.²

LONG-TERM FOLLOW-UP/LATE COMPLICATIONS

At the completion of treatment, sites originally involved with tumor are reassessed, usually by radiography or imaging techniques, and any persistent abnormality is biopsied. If disease persists, the multidisciplinary team discusses a new salvage treatment plan. If the patient has been rendered disease-free by the original treatment, the patient is followed regularly for disease recurrence. The optimal guidelines for follow-up care are not known. For many years, a routine practice has been to follow the patient monthly for 6–12 months, then every other month for a year, every 3 months for a year, every 4 months for a year, every 6 months for a year, and then annually. At each visit, a battery of laboratory and radiographic and imaging tests was obtained on the assumption that it is best to detect recurrent disease before it becomes symptomatic. However, where follow-up procedures have been examined, this assumption has been found to be untrue. Studies of breast cancer, melanoma, lung cancer, colon cancer, and lymphoma have all failed to support the notion that asymptomatic relapses are more readily cured by salvage therapy than symptomatic relapses. In view of the enormous cost of a full battery of diagnostic tests and their manifest lack of impact on survival, new guidelines are emerging for less frequent follow-up visits, during which the history and physical examination are the major investigations performed.

As time passes, the likelihood of recurrence of the primary cancer diminishes. For many types of cancer, survival for 5 years without recurrence is tantamount to cure. However, important medical problems can occur in patients treated for cancer and must be examined (*Chap. 95*). Some problems emerge as a consequence of the disease and some as a consequence of the treatment. An understanding of these disease- and treatment-related problems may help in their detection and management.

Despite these concerns, most patients who are cured of cancer return to normal lives.

SUPPORTIVE CARE

In many ways, the success of cancer therapy depends on the success of the supportive care. Failure to control the symptoms of cancer and its treatment may lead patients to abandon curative therapy. Of equal importance, supportive care is a major determinant of quality of life. Even when life cannot be prolonged, the physician must strive to preserve its quality. Quality-of-life measurements have become common endpoints of clinical research studies. Furthermore, palliative care has been shown to be cost-effective when approached in an organized fashion. A credo for oncology could be to cure sometimes, to extend life often, and to comfort always.

Pain Pain occurs with variable frequency in the cancer patient: 25–50% of patients present with pain at diagnosis, 33% have pain associated with treatment, and 75% have pain with progressive disease. The pain may have several causes. In ~70% of cases, pain is caused by the tumor itself—by invasion of bone, nerves, blood vessels, or mucous membranes or obstruction of a hollow viscus or duct. In ~20% of cases, pain is related to a surgical or invasive medical procedure, to radiation injury (mucositis, enteritis, or plexus, or spinal cord injury), or to chemotherapy injury (mucositis, peripheral neuropathy, phlebitis,

steroid-induced aseptic necrosis of the femoral head). In 10% of cases, pain is unrelated to cancer or its treatment.

Assessment of pain requires the methodical investigation of the history of the pain, its location, character, temporal features, provocative and palliative factors, and intensity (*Chaps. 12 and 13*); a review of the oncologic history and past medical history as well as personal and social history; and a thorough physical examination. The patient should be given a 10-division visual analogue scale on which to indicate the severity of the pain. The clinical condition is often dynamic, making it necessary to reassess the patient frequently. Pain therapy should not be withheld while the cause of pain is being sought.

A variety of tools are available with which to address cancer pain. About 85% of patients will have pain relief from pharmacologic intervention. However, other modalities, including antitumor therapy (such as surgical relief of obstruction, radiation therapy, and strontium-89 or samarium-153 treatment for bone pain), neurostimulatory techniques, regional analgesia, or neuroablative procedures, are effective in an additional 12% or so. Thus, very few patients will have inadequate pain relief if appropriate measures are taken. **A specific approach to pain relief is detailed in Chap. 12.**

Nausea Emesis in the cancer patient is usually caused by chemotherapy (*Chap. 73*). Its severity can be predicted from the drugs used to treat the cancer. Three forms of emesis are recognized on the basis of their timing with regard to the noxious insult. *Acute emesis*, the most common variety, occurs within 24 h of treatment. *Delayed emesis* occurs 1–7 days after treatment; it is rare, but, when present, usually follows cisplatin administration. *Anticipatory emesis* occurs before administration of chemotherapy and represents a conditioned response to visual and olfactory stimuli previously associated with chemotherapy delivery.

Acute emesis is the best understood form. Stimuli that activate signals in the chemoreceptor trigger zone in the medulla, the cerebral cortex, and peripherally in the intestinal tract lead to stimulation of the vomiting center in the medulla, the motor center responsible for coordinating the secretory and muscle contraction activity that leads to emesis. Diverse receptor types participate in the process, including dopamine, serotonin, histamine, opioid, and acetylcholine receptors. The serotonin receptor antagonists ondansetron and granisetron are effective drugs against highly emetogenic agents, as are neurokinin receptor antagonists such as aprepitant and fosaprepitant (see *Chap. 73*).

As with the analgesia ladder, emesis therapy should be tailored to the situation. For mildly and moderately emetogenic agents, prochlorperazine, 5–10 mg PO or 25 mg PR, is effective. Its efficacy may be enhanced by administering the drug before the chemotherapy is delivered. Dexamethasone, 10–20 mg IV, is also effective and may enhance the efficacy of prochlorperazine. For highly emetogenic agents such as cisplatin, mechlorethamine, dacarbazine, and streptozocin, combinations of agents work best and administration should begin 6–24 h before treatment. Ondansetron, 8 mg PO every 6 h the day before therapy and IV on the day of therapy, plus dexamethasone, 20 mg IV before treatment, is an effective regimen. Addition of oral aprepitant (a substance P/neurokinin 1 receptor antagonist) to this regimen (125 mg on day 1, 80 mg on days 2 and 3) further decreases the risk of both acute and delayed vomiting. Like pain, emesis is easier to prevent than to alleviate.

Delayed emesis may be related to bowel inflammation from the therapy and can be controlled with oral dexamethasone and oral metoclopramide, a dopamine receptor antagonist that also blocks serotonin receptors at high dosages. The best strategy for preventing anticipatory emesis is to control emesis in the early cycles of therapy to prevent the conditioning from taking place. If this is unsuccessful, prophylactic antiemetics the day before treatment may help. Experimental studies are evaluating behavior modification.

Effusions Fluid may accumulate abnormally in the pleural cavity, pericardium, or peritoneum. Asymptomatic malignant effusions may not require treatment. Symptomatic effusions occurring in tumors responsive to systemic therapy usually do not require local treatment

²Information about unsound methods may be obtained from the National Council Against Health Fraud, Box 1276, Loma Linda, CA 92354, or from the Center for Medical Consumers and Health Care Information, 237 Thompson Street, New York, NY 10012.

but respond to the treatment for the underlying tumor. Symptomatic effusions occurring in tumors unresponsive to systemic therapy may require local treatment in patients with a life expectancy of at least 6 months.

Pleural effusions due to tumors may or may not contain malignant cells. Lung cancer, breast cancer, and lymphomas account for ~75% of malignant pleural effusions. Their exudative nature is usually gauged by an effusion/serum protein ratio of ≥ 0.5 or an effusion/serum lactate dehydrogenase ratio of ≥ 0.6 . When the condition is symptomatic, thoracentesis is usually performed first. In most cases, symptomatic improvement occurs for <1 month. Chest tube drainage is required if symptoms recur within 2 weeks. Fluid is aspirated until the flow rate is <100 mL in 24 h. Then either 60 units of bleomycin or 1 g of doxycycline is infused into the chest tube in 50 mL of 5% dextrose in water; the tube is clamped; the patient is rotated on four sides, spending 15 min in each position; and, after 1–2 h, the tube is again attached to suction for another 24 h. The tube is then disconnected from suction and allowed to drain by gravity. If <100 mL drains over the next 24 h, the chest tube is pulled, and a radiograph is taken 24 h later. If the chest tube continues to drain fluid at an unacceptably high rate, sclerosis can be repeated. Bleomycin may be somewhat more effective than doxycycline but is very expensive. Doxycycline is usually the drug of first choice. If neither doxycycline nor bleomycin is effective, talc can be used.

Symptomatic pericardial effusions are usually treated by creating a pericardial window or by stripping the pericardium. If the patient's condition does not permit a surgical procedure, sclerosis can be attempted with doxycycline and/or bleomycin.

Malignant ascites is usually treated with repeated paracentesis of small volumes of fluid. If the underlying malignancy is unresponsive to systemic therapy, peritoneovenous shunts may be inserted. Despite the fear of disseminating tumor cells into the circulation, widespread metastases are an unusual complication. The major complications are occlusion, leakage, and fluid overload. Patients with severe liver disease may develop disseminated intravascular coagulation.

Nutrition Cancer and its treatment may lead to a decrease in nutrient intake of sufficient magnitude to cause weight loss and alteration of intermediary metabolism. The prevalence of this problem is difficult to estimate because of variations in the definition of cancer cachexia, but most patients with advanced cancer experience weight loss and decreased appetite. A variety of both tumor-derived factors (e.g., bombesin, adrenocorticotrophic hormone) and host-derived factors (e.g., tumor necrosis factor, interleukins 1 and 6, growth hormone) contribute to the altered metabolism, and a vicious cycle is established in which protein catabolism, glucose intolerance, and lipolysis cannot be reversed by the provision of calories.

It remains controversial how to assess nutritional status and when and how to intervene. Efforts to make the assessment objective have included the use of a prognostic nutritional index based on albumin levels, triceps skinfold thickness, transferrin levels, and delayed-type hypersensitivity skin testing. However, a simpler approach has been to define the threshold for nutritional intervention as <10% unexplained body weight loss, serum transferrin level <1500 mg/L (150 mg/dL), and serum albumin <34 g/L (3.4 g/dL).

The decision is important, because it appears that cancer therapy is substantially more toxic and less effective in the face of malnutrition. Nevertheless, it remains unclear whether nutritional intervention can alter the natural history. Unless some pathology is affecting the absorptive function of the gastrointestinal tract, enteral nutrition provided orally or by tube feeding is preferred over parenteral supplementation. However, the risks associated with the tube may outweigh the benefits. Megestrol acetate, a progestational agent, has been advocated as a pharmacologic intervention to improve nutritional status. Research in this area may provide more tools in the future as cytokine-mediated mechanisms are further elucidated.

Psychosocial Support The psychosocial needs of patients vary with their situation. Patients undergoing treatment experience fear, anxiety, and depression. Self-image is often seriously compromised

by deforming surgery and loss of hair. Women who receive cosmetic advice that enables them to look better also feel better. Loss of control over how one spends time can contribute to the sense of vulnerability. Juggling the demands of work and family with the demands of treatment may create enormous stresses. Sexual dysfunction is highly prevalent and needs to be discussed openly with the patient. An empathetic health care team is sensitive to the individual patient's needs and permits negotiation where such flexibility will not adversely affect the course of treatment.

Cancer survivors have other sets of difficulties. Patients may have fears associated with the termination of a treatment they associate with their continued survival. Adjustments are required to physical losses and handicaps, real and perceived. Patients may be preoccupied with minor physical problems. They perceive a decline in their job mobility and view themselves as less desirable workers. They may be victims of job and/or insurance discrimination. Patients may experience difficulty reentering their normal past life. They may feel guilty for having survived and may carry a sense of vulnerability to colds and other illnesses. Perhaps the most pervasive and threatening concern is the ever-present fear of relapse (the Damocles syndrome).

Patients in whom therapy has been unsuccessful have other problems related to the end of life.

Death and Dying The most common causes of death in patients with cancer are infection (leading to circulatory failure), respiratory failure, hepatic failure, and renal failure. Intestinal blockage may lead to inanition and starvation. Central nervous system disease may lead to seizures, coma, and central hypoventilation. About 70% of patients develop dyspnea preterminally. However, many months usually pass between the diagnosis of cancer and the occurrence of these complications, and during this period, the patient is severely affected by the possibility of death. The path of unsuccessful cancer treatment usually occurs in three phases. First, there is optimism at the hope of cure; when the tumor recurs, there is the acknowledgment of an incurable disease, and the goal of palliative therapy is embraced in the hope of being able to live with disease; finally, at the disclosure of imminent death, another adjustment in outlook takes place. The patient imagines the worst in preparation for the end of life and may go through stages of adjustment to the diagnosis. These stages include denial, isolation, anger, bargaining, depression, acceptance, and hope. Of course, patients do not all progress through all the stages or proceed through them in the same order or at the same rate. Nevertheless, developing an understanding of how the patient has been affected by the diagnosis and is coping with it is an important goal of patient management.

It is best to speak frankly with the patient and the family regarding the likely course of disease. These discussions can be difficult for the physician as well as for the patient and family. The critical features of the interaction are to reassure the patient and family that everything that can be done to provide comfort will be done. They will not be abandoned. Many patients prefer to be cared for in their homes or in a hospice setting rather than a hospital. The American College of Physicians has published a book called *Home Care Guide for Cancer: How to Care for Family and Friends at Home* that teaches an approach to successful problem-solving in home care. With appropriate planning, it should be possible to provide the patient with the necessary medical care as well as the psychological and spiritual support that will prevent the isolation and depersonalization that can attend in-hospital death.

The care of dying patients may take a toll on the physician. A "burn-out" syndrome has been described that is characterized by fatigue, disengagement from patients and colleagues, and a loss of self-fulfillment. Efforts at stress reduction, maintenance of a balanced life, and setting realistic goals may combat this disorder.

End-of-Life Decisions Unfortunately, a smooth transition in treatment goals from curative to palliative may not be possible in all cases because of the occurrence of serious treatment-related complications or rapid disease progression. Vigorous and invasive medical support for a reversible disease or treatment complication is assumed to be justified. However, if the reversibility of the condition is in doubt,

the patient's wishes determine the level of medical care. These wishes should be elicited before the terminal phase of illness and reviewed periodically. Information about advance directives can be obtained from the American Association of Retired Persons, 601 E Street, NW, Washington, DC 20049, 202-434-2277, or Choice in Dying, 250 West 57th Street, New York, NY 10107, 212-366-5540. Some states allow physicians to assist patients who choose to end their lives. This subject is challenging from an ethical and a medical point of view. Discussions of end-of-life decisions should be candid and involve clear informed consent, waiting periods, second opinions, and documentation. **A full discussion of end-of-life management is provided in Chap. 12.**

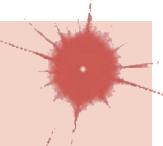
FURTHER READING

- BRAY F et al: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394, 2018.
- HESKETH PJ et al: Antiemetics: ASCO guideline update. J Clin Oncol 38:2782, 2020.
- KELLEY AS, MORRISON RS: Palliative care for the seriously ill. N Engl J Med 373:747, 2015.
- SAMALA RV et al: Frequently asked questions about managing cancer pain: An update. Cleve Clin Med J 88:183, 2021.
- SIEGEL RL et al: Cancer statistics, 2021. CA Cancer J Clin 71:7, 2021.

70

Prevention and Early Detection of Cancer

Jennifer M. Croswell, Otis W. Brawley,
Barnett S. Kramer



Improved understanding of carcinogenesis has allowed cancer prevention and early detection to expand beyond identification and avoidance of carcinogens. Specific interventions to reduce cancer mortality by preventing cancer in those at risk and effective screening for early detection of cancer are the goals.

Carcinogenesis is a process that usually extends over years, a continuum of discrete tissue and cellular changes over time resulting in aberrant physiologic processes. Prevention concerns the identification and manipulation of the biologic, environmental, social, and genetic factors in the causal pathway of cancer. Examination of national epidemiologic patterns can provide indicators of the relative contributions of advances in prevention, screening, and therapy in progress against cancer, but randomized trials provide the best evidence to guide practice, especially in the healthy general population.

EDUCATION AND HEALTHFUL HABITS

Public education on the avoidance of identified risk factors for cancer and encouraging healthy habits contributes to cancer prevention. The clinician is a powerful messenger in this process. The patient-provider encounter provides an opportunity to teach patients about the hazards of smoking, influence of a healthy lifestyle and other exposures, and use of proven cancer screening methods.

SMOKING CESSATION

Tobacco smoking is a strong, modifiable risk factor for cardiovascular disease, pulmonary disease, and cancer. Smokers have an ~1 in 3 lifetime risk of dying prematurely from a tobacco-related cancer, cardiovascular, or pulmonary disease. Tobacco use causes more deaths from cardiovascular disease than from cancer. Lung cancer and cancers of the larynx, oropharynx, esophagus, kidney, bladder, colon, pancreas, stomach, and uterine cervix are all tobacco related.

The number of cigarettes smoked per day and the level of inhalation of cigarette smoke are correlated with risk of lung cancer mortality. Light- and low-tar cigarettes are not safer because smokers tend to inhale them more frequently and deeply.

Those who stop smoking have a 30–50% lower 10-year lung cancer mortality rate compared to those who continue smoking, despite the fact that some carcinogen-induced gene mutations persist for years after smoking cessation. Smoking cessation and avoidance would save more lives than any other public health activity.

The risk of tobacco smoke is not limited to the smoker. Environmental tobacco smoke, known as secondhand or passive smoke, is carcinogenic and associated with a variety of respiratory illnesses in exposed children.

Tobacco use prevention is a pediatric issue. More than 80% of adult American smokers began smoking before the age of 18 years. Cigarette smoking has been declining in recent years, but in recent surveys, about 8% of high school students reported smoking within the prior month. Electronic cigarettes, on the other hand, are rapidly increasing in use: approximately 28% of high school students and 11% of middle school students are current electronic cigarette users. Counseling of adolescents and young adults is critical to prevent all forms of tobacco use. A clinician's simple advice can be of benefit. Providers should query patients on tobacco use and offer smokers assistance in quitting.

Current approaches to smoking cessation recognize nicotine in tobacco as addicting (**Chap. 454**). The smoker who is quitting goes through identifiable stages, including contemplation of quitting, an action phase in which the smoker quits, and a maintenance phase. Smokers who quit completely are more likely to be successful than those who gradually reduce the number of cigarettes smoked or change to lower-tar or lower-nicotine cigarettes. Organized cessation programs may help individual efforts. Heavy smokers may need an intensive broad-based cessation program that includes counseling, behavioral strategies, and pharmacologic adjuncts, such as nicotine replacement (gum, patches, sprays, lozenges, and inhalers), bupropion, and/or varenicline. Electronic cigarettes have been advocated as a tool to achieve smoking cessation in adults, but it is not known how effective electronic cigarettes are for this purpose. The net effects of electronic cigarettes on health are poorly studied. Absence of strict manufacturing controls of vaping material has produced serious injury.

The health risks of cigars are similar to those of cigarettes. Smoking one or two cigars daily doubles the risk for oral and esophageal cancers; smoking three or four cigars daily increases the risk of oral cancers more than eightfold and esophageal cancer fourfold. The risks of occasional use are unknown.

Smokeless tobacco also represents a substantial health risk. Chewing tobacco is a carcinogen linked to dental caries, gingivitis, oral leukoplakia, and oral cancer. The systemic effects of smokeless tobacco (including snuff) may increase risks for other cancers. Esophageal cancer is linked to carcinogens in tobacco dissolved in saliva and swallowed. The net effects of e-cigarettes on health are poorly studied.

PHYSICAL ACTIVITY

Physical activity is associated with a decreased risk of colon and breast cancer. A variety of mechanisms have been proposed. However, such studies are prone to confounding factors such as recall bias, association of exercise with other health-related practices, and effects of preclinical cancers on exercise habits (reverse causality).

DIET MODIFICATION

International epidemiologic studies suggest that diets high in fat are associated with increased risk for cancers of the breast, colon, prostate, and endometrium. Despite correlations, dietary fat has not been proven to cause cancer. Case-control and cohort epidemiologic studies give conflicting results. Diet is a highly complex exposure to many nutrients and chemicals. Low-fat diets are associated with many dietary changes beyond simple subtraction of fat. Other lifestyle factors are also associated with adherence to a low-fat diet.

In some observational studies, dietary fiber has been associated with a reduced risk of colonic polyps and invasive cancer of the colon.

Two large prospective cohort studies of >100,000 health professionals showed no association between fruit and vegetable intake and risk of cancer, however. Cancer-protective effects of increasing fiber and lowering dietary fat have not been shown in the context of a prospective clinical trial. The Polyp Prevention Trial randomly assigned 2000 elderly persons, who had polyps removed, to a low-fat, high-fiber diet versus routine diet for 4 years. No differences were noted in polyp formation.

The U.S. National Institutes of Health Women's Health Initiative, launched in 1994, was a long-term clinical trial enrolling >100,000 women age 45–69 years. It placed women into 22 intervention groups. Participants received calcium/vitamin D supplementation; hormone replacement therapy; and counseling to increase exercise, eat a low-fat diet with increased consumption of fruits, vegetables, and fiber, and cease smoking. The study showed that although dietary fat intake was lower in the diet intervention group, invasive breast cancers were not reduced over an 8-year follow-up period compared to the control group. Additionally, no reduction was seen in the incidence of colorectal cancer in the dietary intervention arm. In the aggregate, cohort studies and randomized trials suggest that reduction of red meat or processed meat consumption has a small (if any) effect on cancer incidence and mortality, although the overall evidence base is weak. Evidence does not currently establish the anticarcinogenic value of vitamin, mineral, or nutritional supplements in amounts greater than those provided by a balanced diet.

■ ENERGY BALANCE

Risk of certain cancers appears to increase modestly (relative risks generally in the 1.0–2.0 range) as body mass index (BMI) increases beyond 25 kg/m². A cohort study of >5 million adults included in the U.K. Clinical Practice Research Datalink (a primary care database) found that each 5 kg/m² increase in BMI was linearly associated with cancers of the uterus, gallbladder, kidney, cervix, thyroid, and leukemia. High BMI appears to have an inverse association with prostate and premenopausal breast cancer.

■ SUN AVOIDANCE

Nonmelanoma skin cancers (basal cell and squamous cell) are induced by cumulative exposure to ultraviolet (UV) radiation. Sunburns, especially in childhood and adolescence, may be associated with an increased risk of melanoma in adulthood. Reduction of sun exposure through use of protective clothing and changing patterns of outdoor activities can reduce skin cancer risk. Sunscreens decrease the risk of actinic keratoses, the precursor to squamous cell skin cancer, but melanoma risk may not be reduced. Sunscreens prevent burning, but they may encourage more prolonged exposure to the sun and may not filter out wavelengths of energy that cause melanoma.

Appearance-focused behavioral interventions in young women can decrease indoor tanning use and other UV exposures and may be more effective than messages about long-term cancer risks. Those who recognize themselves as being at risk tend to be more compliant with sun-avoidance recommendations. Risk factors for melanoma include a propensity to sunburn, a large number of benign melanocytic nevi, and atypical nevi.

CANCER CHEMOPREVENTION

Chemoprevention involves the use of specific natural or synthetic chemical agents to reverse, suppress, or prevent carcinogenesis before the development of invasive malignancy.

Cancer develops through an accumulation of tissue abnormalities associated with genetic and epigenetic changes, and growth regulatory pathways that are potential points of intervention to prevent cancer. The initial changes are termed *initiation*. The alteration can be inherited or acquired through the action of physical, infectious, or chemical carcinogens. Like most human diseases, cancer arises from an interaction between genetics and environmental exposures (Table 70-1). Influences that cause the initiated cell and its surrounding tissue microenvironment to progress through the carcinogenic process and change phenotypically are termed *promoters*. Promoters

TABLE 70-1 Suspected Carcinogens

CARCINOGENS*	ASSOCIATED CANCER OR NEOPLASM
Aalkylating agents	Acute myeloid leukemia, bladder cancer
Androgens	Prostate cancer
Aromatic amines (dyes)	Bladder cancer
Arsenic	Cancer of the lung, skin
Asbestos	Cancer of the lung, pleura, peritoneum
Benzene	Acute myelocytic leukemia
Chromium	Lung cancer
Diethylstilbestrol (prenatal)	Vaginal cancer (clear cell)
Epstein-Barr virus	Burkitt's lymphoma, nasal T-cell lymphoma
Estrogens	Cancer of the endometrium, liver, breast
Ethyl alcohol	Cancer of the breast, liver, esophagus, head and neck
<i>Helicobacter pylori</i>	Gastric cancer, gastric mucosa-associated lymphoid tissue (MALT) lymphoma
Hepatitis B or C virus	Liver cancer
Human immunodeficiency virus	Non-Hodgkin's lymphoma, Kaposi's sarcoma, squamous cell carcinomas (especially of the urogenital tract)
Human papillomavirus	Cancers of the cervix, anus, oropharynx
Human T-cell lymphotropic virus type 1 (HTLV-1)	Adult T-cell leukemia/lymphoma
Immunosuppressive agents (azathioprine, cyclosporine, glucocorticoids)	Non-Hodgkin's lymphoma
Ionizing radiation (therapeutic or diagnostic)	Breast, bladder, thyroid, soft tissue, bone, hematopoietic, and many more
Nitrogen mustard gas	Cancer of the lung, head and neck, nasal sinuses
Nickel dust	Cancer of the lung, nasal sinuses
Diesel exhaust	Lung cancer (miners)
Phenacetin	Cancer of the renal pelvis and bladder
Polycyclic hydrocarbons	Cancer of the lung, skin (especially squamous cell carcinoma of scrotal skin)
Radon gas	Lung cancer
Schistosomiasis	Bladder cancer (squamous cell)
Sunlight (ultraviolet)	Skin cancer (squamous cell and melanoma)
Tobacco (including smokeless)	Cancer of the upper aerodigestive tract, bladder
Vinyl chloride	Liver cancer (angiosarcoma)

*Agents that are thought to act as cancer initiators and/or promoters.

include hormones such as androgens, linked to prostate cancer, and estrogen, linked to breast and endometrial cancer. The difference between an initiator and promoter is indistinct; some components of cigarette smoke are “complete carcinogens,” acting as both initiators and promoters. Cancer can be prevented or controlled through interference with the factors that cause cancer initiation, promotion, or progression. Compounds of interest in chemoprevention often have antimutagenic, hormone modulation, anti-inflammatory, antiproliferative, or proapoptotic activity (or a combination).

■ CHEMOPREVENTION OF CANCERS OF THE UPPER AERODIGESTIVE TRACT

Smoking causes diffuse epithelial injury in the oral cavity, neck, esophagus, and lung. Patients cured of squamous cell cancers of the lung, esophagus, oral cavity, and neck are at risk (as high as 5% per year) of developing second cancers of the upper aerodigestive tract. Cessation of cigarette smoking does not markedly decrease the cured cancer patient's risk of second malignancy, even though it does lower the cancer risk in those who have never developed a malignancy. Smoking cessation may halt the early stages of the carcinogenic process (such as metaplasia), but it may have no effect on late stages of carcinogenesis.

This “field carcinogenesis” hypothesis for upper aerodigestive tract cancer has made “cured” patients an important population for chemoprevention of second malignancies.

Persistent oral human papillomavirus (HPV) infection, particularly HPV-16, increases the risk for cancers of the oropharynx. This association exists even in the absence of other risk factors such as smoking or alcohol use (although the magnitude of increased risk appears greater than additive when HPV infection and smoking are both present). Oral HPV infection is believed to be largely sexually acquired. Although the evidence is not definitive, the use of the HPV vaccine is associated with a reduction in prevalence of oropharyngeal infection rates and may eventually reduce oropharyngeal cancer rates (unlike cancers of the cervix, no precursor lesion for oropharyngeal tumors is known).

Oral leukoplakia, a premalignant lesion commonly found in smokers, has been used as an intermediate marker of chemopreventive activity in smaller shorter-duration, randomized, placebo-controlled trials. Although therapy with high, relatively toxic doses of isotretinoin (*13-cis*-retinoic acid) causes regression of oral leukoplakia, more tolerable doses of isotretinoin have not shown benefit in the prevention of head and neck cancer.

Several large-scale trials have assessed agents in the chemoprevention of lung cancer in patients at high risk. In the α -tocopherol/ β -carotene (ATBC) Lung Cancer Prevention Trial, participants were male smokers, age 50–69 years at entry. Participants had smoked an average of one pack of cigarettes per day for nearly 36 years. Participants received α -tocopherol, β -carotene, and/or placebo in a randomized, two-by-two factorial design. After median follow-up of 6 years, lung cancer incidence and mortality were statistically significantly increased in those receiving β -carotene. α -Tocopherol had no effect on lung cancer mortality. However, patients receiving α -tocopherol had a higher incidence of hemorrhagic stroke.

The β -Carotene and Retinol Efficacy Trial (CARET) involved 17,000 American smokers and workers with asbestos exposure. Entrants were randomly assigned to one of four arms and received β -carotene, retinol, and/or placebo in a two-by-two factorial design. This trial also demonstrated harm from β -carotene: a lung cancer rate of 5 per 1000 subjects per year for those taking placebo versus 6 per 1000 subjects per year for those taking β -carotene.

The ATBC and CARET results demonstrate the importance of testing chemoprevention hypotheses thoroughly before widespread implementation because the results contradict a number of observational studies.

CHEMOPREVENTION OF COLON CANCER

Many colon cancer prevention trials are based on the premise that most colorectal cancers develop from adenomatous polyps. These trials use adenoma recurrence or disappearance as a surrogate endpoint (not yet validated) for colon cancer prevention. Clinical trial results suggest that nonsteroidal anti-inflammatory drugs (NSAIDs), such as piroxicam, sulindac, and aspirin, may prevent adenoma formation or cause regression of adenomatous polyps. The mechanism of action of NSAIDs is unknown, but they are presumed to work through the cyclooxygenase pathway. A meta-analysis of four randomized controlled trials (albeit primarily designed to examine aspirin’s effects on cardiovascular events) found that aspirin at doses of at least 75 mg/d resulted in a 33% relative reduction in colorectal cancer incidence after 20 years, with no clear increase in efficacy at higher doses. Based on a systematic review of evidence from randomized trials for primary prevention of cardiovascular disease, the U.S. Preventive Services Task Force concluded that the balance of benefits and harms favored initiating low-dose aspirin for colorectal cancer and cardiovascular disease prevention in adults age 50–59 if they have a 10% or greater 10-year risk of cardiovascular disease. Low-dose aspirin does not appear to benefit the elderly, however. The ASPREE trial, which compared 100 mg of daily aspirin to placebo for improvement in the composite endpoint of death, dementia, or survival in the healthy elderly, was stopped because of a lack of benefit, including cancer. Cyclooxygenase-2 (COX-2) inhibitors have been considered for colorectal cancer and polyp prevention. Trials with COX-2 inhibitors were initiated, but an increased risk of cardiovascular events

in those taking the COX-2 inhibitors was noted, suggesting that these agents are not suitable for chemoprevention in the general population.

The Women’s Health Initiative demonstrated that postmenopausal women taking estrogen plus progestin have a 44% lower relative risk of colorectal cancer compared to women taking placebo. Of >16,600 women randomized and followed for a median of 5.6 years, 43 invasive colorectal cancers occurred in the hormone group and 72 in the placebo group. The positive effect on colon cancer is mitigated by the modest increase in cardiovascular and breast cancer risks associated with combined estrogen plus progestin therapy.

Most case-control and cohort studies have not confirmed early reports of an association between regular statin use and a reduced risk of colorectal cancer. No randomized controlled trials have addressed this hypothesis. A meta-analysis of statin use showed no protective effect of statins on overall cancer incidence or death.

CHEMOPREVENTION OF BREAST CANCER

Tamoxifen is an antiestrogen with partial estrogen agonistic activity in some tissues, such as endometrium and bone. One of its actions is to upregulate transforming growth factor β , which decreases breast cell proliferation. In a randomized placebo-controlled prevention trial involving >13,000 pre- and postmenopausal women at high risk, tamoxifen decreased the risk of developing breast cancer by 49% (from 43.4 to 22 per 1000 women) after a median follow-up of nearly 6 years. Tamoxifen also reduced bone fractures; a small increase in risk of endometrial cancer, stroke, pulmonary emboli, and deep vein thrombosis was noted. The International Breast Cancer Intervention Study (IBIS-I) and the Italian Randomized Tamoxifen Prevention Trial also demonstrated a reduction in breast cancer incidence with tamoxifen use. A trial comparing tamoxifen with another selective estrogen receptor modulator, raloxifene, performed in postmenopausal women showed that raloxifene is comparable to tamoxifen in cancer prevention, but without the risk of endometrial cancer. Raloxifene was associated with a smaller reduction in invasive breast cancers and a trend toward more noninvasive breast cancers, but fewer thromboembolic events than tamoxifen; the drugs are similar in risks of other cancers, fractures, ischemic heart disease, and stroke. Both tamoxifen and raloxifene (the latter for postmenopausal women only) have been approved by the U.S. Food and Drug Administration (FDA) for reduction of breast cancer in women at high risk for the disease (1.66% risk at 5 years based on the Gail risk model: <http://www.cancer.gov/bcrisktool/>).

Because the aromatase inhibitors are even more effective than tamoxifen in adjuvant breast cancer therapy, it has been hypothesized that they would be more effective in breast cancer prevention. A randomized, placebo-controlled trial of exemestane reported a 65% relative reduction (from 5.5 to 1.9 per 1000 women) in the incidence of invasive breast cancer in women at elevated risk after a median follow-up of about 3 years. Common adverse effects included arthralgias, hot flashes, fatigue, and insomnia. No trial has directly compared aromatase inhibitors with selective estrogen receptor modulators for breast cancer chemoprevention.

CHEMOPREVENTION OF PROSTATE CANCER

Finasteride and dutasteride are 5- α -reductase inhibitors. They inhibit conversion of testosterone to dihydrotestosterone (DHT), a potent stimulator of prostate cell proliferation. The Prostate Cancer Prevention Trial (PCPT) randomly assigned men age 55 years or older at average risk of prostate cancer to finasteride or placebo. All men in the trial were being regularly screened with prostate-specific antigen (PSA) levels and digital rectal examination. After 7 years of therapy, the incidence of prostate cancer was 18.4% in the finasteride arm, compared with 24.4% in the placebo arm, a statistically significant difference. However, the finasteride group had more patients with tumors of Gleason score 7 and higher compared with the placebo arm (6.4 vs 5.1%). Long-term (10–15 years) follow-up did not reveal any statistically significant differences in overall or prostate cancer-specific mortality between all men in the finasteride and placebo arms or in men diagnosed with prostate cancer, but the power to detect a difference was limited.

Dutasteride has also been evaluated as a preventive agent for prostate cancer. The Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial was a randomized double-blind trial in which ~8200 men with an elevated PSA (2.5–10 ng/mL for men age 50–60 years and 3–10 ng/mL for men age 60 years or older) and negative prostate biopsy on enrollment received daily 0.5 mg of dutasteride or placebo. The trial found a statistically significant 23% relative risk reduction in the incidence of biopsy-detected prostate cancer in the dutasteride arm at 4 years of treatment (659 cases vs 858 cases, respectively). Overall, across years 1 through 4, no difference was seen between the arms in the number of tumors with a Gleason score of 7 to 10; however, during years 3 and 4, there was a statistically significant difference in tumors with Gleason score of 8 to 10 in the dutasteride arm (12 tumors vs 1 tumor, respectively).

The clinical importance of the apparent increased incidence of higher-grade tumors in the 5- α -reductase inhibitor arms of these trials likely represents an increased sensitivity of PSA and digital rectal exam for high-grade tumors in men receiving these agents due to a decrease in prostatic volume. Although the FDA acknowledged that detection bias may have accounted for the finding, a causative role for 5- α -reductase inhibitors could not be conclusively dismissed. These agents are therefore not FDA-approved for prostate cancer prevention.

Because all men in both the PCPT and REDUCE trials were being screened and because screening approximately doubles the rate of prostate cancer, it is not known if finasteride or dutasteride decreases the risk of prostate cancer in men who are not being screened or simply reduces the risk of non-life-threatening cancers detectable by screening.

Several favorable laboratory and observational studies led to the formal evaluation of selenium and α -tocopherol (vitamin E) as potential prostate cancer preventives. The Selenium and Vitamin E Cancer Prevention Trial (SELECT) assigned 35,533 men to receive 200 μ g/d selenium, 400 IU/d α -tocopherol, selenium plus vitamin E, or placebo. After a median follow-up of 7 years, a trend toward an increased risk of developing prostate cancer was observed for men taking vitamin E alone as compared to the placebo arm (hazard ratio 1.17; 95% confidence interval, 1.004–1.36).

VACCINES AND CANCER PREVENTION

A number of infectious agents cause cancer. Hepatitis B and C are linked to liver cancer; some HPV strains are linked to cervical, anal, and head and neck cancer; and *Helicobacter pylori* is associated with gastric adenocarcinoma and gastric lymphoma. Vaccines to protect against these agents may therefore reduce the risk of their associated cancers.

The hepatitis B vaccine is effective in preventing hepatitis and hepatomas due to chronic hepatitis B infection.

A nonavalent vaccine (covering HPV strains 6, 11, 16, 18, 31, 33, 45, 52, and 58) is available for use in the United States. HPV types 6 and 11 cause genital papillomas. The remaining HPV types cause cervical and anal cancer; reduction in HPV types 16 and 18 alone could prevent >70% of cervical cancers worldwide. For individuals not previously infected with these HPV strains, the vaccine demonstrates high efficacy in preventing persistent strain-specific HPV infections. Studies also confirm the vaccine's ability to prevent preneoplastic lesions (cervical or anal intraepithelial neoplasia [CIN/AIN] I, II, and III). The durability of the immune response beyond 10–12 years is not currently known. The vaccine does not appear to impact preexisting infections. A two-dose schedule is currently recommended in the United States for females and males age 9–14 years; teens and young adults who start the series between 15 and 26 years are recommended to receive three doses of the vaccine. However, observational studies suggest similar efficacy with a single dose in young girls, and a large randomized trial is currently comparing one to two doses.

SURGICAL PREVENTION OF CANCER

Some organs in some individuals are at such high risk of developing cancer that surgical removal of the organ at risk may be considered. Women with severe cervical dysplasia are treated with laser or loop

electrosurgical excision or conization. Colectomy may be used to prevent colon cancer in patients with familial polyposis or ulcerative colitis.

Prophylactic bilateral mastectomy may be chosen for breast cancer prevention among women with genetic predisposition to breast cancer. In a prospective series of 139 women with *BRCA1* and *BRCA2* mutations, 76 chose to undergo prophylactic mastectomy, and 63 chose close surveillance. At 3 years, no cases of breast cancer had been diagnosed in those opting for surgery, but eight patients in the surveillance group had developed breast cancer. A larger ($n = 639$) retrospective cohort study reported that three patients developed breast cancer after prophylactic mastectomy compared with an expected incidence of 30–53 cases. Postmastectomy breast cancer-related deaths were 81–94% lower in high-risk women compared with sister controls and 100% lower in moderate-risk women when compared with expected rates.

Prophylactic salpingo-oophorectomy may also be employed for the prevention of ovarian and breast cancers among high-risk women. A prospective cohort study evaluating the outcomes of *BRCA* mutation carriers demonstrated a statistically significant association between prophylactic salpingo-oophorectomy and a reduced incidence of ovarian or primary peritoneal cancer (36% relative risk reduction, or a 4.5% absolute difference). Studies of prophylactic oophorectomy for prevention of breast cancer in women with genetic mutations have shown relative risks of approximately 0.50; the risk reduction may be greatest for women having the procedure at younger (i.e., <50 years) ages. The observation that most high-grade serous "ovarian cancers" actually arise in the fallopian tube fimbria raises the possibility that this lethal subtype may be prevented by ovary-sparing salpingectomy.

All of the evidence concerning the use of prophylactic mastectomy and salpingo-oophorectomy for prevention of breast and ovarian cancer in high-risk women has been observational in nature; such studies are prone to a variety of biases, including case selection bias, family relationships between patients and controls, and inadequate information about hormone use. Thus, they may give an overestimate of the magnitude of benefit.

CANCER SCREENING

Screening is a means of early detection in asymptomatic individuals, with the goal of decreasing morbidity and mortality. While screening can potentially reduce disease-specific deaths and has been shown to do so in cervical, colon, lung, and breast cancer, it is also subject to a number of biases that can suggest a benefit when actually there is none. Biases can even mask net harm. Early detection does not in itself confer benefit. Cause-specific mortality, rather than survival after diagnosis, is the preferred endpoint (see below).

Because screening is done on asymptomatic, healthy persons, it should offer substantial likelihood of benefit that outweighs harm. Screening tests and their appropriate use should be carefully evaluated before their use is widely encouraged in screening programs.

A large and increasing number of genetic mutations and nucleotide polymorphisms have been associated with an increased risk of cancer. Testing for these genetic mutations could in theory define a high-risk population. However, most of the identified mutations have very low penetrance and individually provide limited predictive accuracy. The ability to predict the development of a particular cancer may someday present therapeutic options as well as ethical dilemmas. It may eventually allow for early intervention to prevent a cancer or limit its severity. People at high risk may be ideal candidates for chemoprevention and screening; however, efficacy of these interventions in the high-risk population should be investigated. Currently, persons at high risk for a particular cancer can engage in intensive screening. While this course is clinically reasonable, it is not known if it reduces mortality in these populations.

The Accuracy of Screening A screening test's accuracy or ability to discriminate disease is described by four indices: sensitivity, specificity, positive predictive value, and negative predictive value (Table 70-2). Sensitivity, also called the true-positive rate, is the proportion of persons with the disease who test positive in the screen

TABLE 70-2 Assessment of the Value of a Diagnostic Test^a

	CONDITION PRESENT	CONDITION ABSENT
Positive test	<i>a</i>	<i>b</i>
Negative test	<i>c</i>	<i>d</i>
<i>a</i> = true positive		
<i>b</i> = false positive		
<i>c</i> = false negative		
<i>d</i> = true negative		
Sensitivity	The proportion of persons with the condition who test positive: $a/(a + c)$	
Specificity	The proportion of persons without the condition who test negative: $d/(b + d)$	
Positive predictive value (PPV)	The proportion of persons with a positive test who have the condition: $a/(a + b)$	
Negative predictive value	The proportion of persons with a negative test who do not have the condition: $d/(c + d)$	
Prevalence, sensitivity, and specificity determine PPV		
PPV = $\frac{\text{prevalence} \times \text{sensitivity}}{(\text{prevalence} \times \text{sensitivity}) + (1 - \text{prevalence})(1 - \text{specificity})}$		

^aFor diseases of low prevalence, such as cancer, poor specificity has a dramatic adverse effect on PPV such that only a small fraction of positive tests are true positives.

(i.e., the ability of the test to detect disease when it is present). *Specificity*, or 1 minus the false-positive rate, is the proportion of persons who do not have the disease who test negative in the screening test (i.e., the ability of a test to correctly indicate that the disease is not present). The *positive predictive value* is the proportion of persons who test positive and who actually have the disease. Similarly, *negative predictive value* is the proportion testing negative who do not have the disease. The sensitivity and specificity of a test are independent of the underlying prevalence (or risk) of the disease in the population screened, but the predictive values depend strongly on the prevalence of the disease.

Screening is most beneficial, efficient, and economical when the target disease is common in the population being screened. Specificity is at least as important to the ultimate feasibility and success of a screening test as sensitivity.

Potential Biases of Screening Tests Common biases of screening are lead time, length-biased sampling, and selection. These biases can make a screening test seem beneficial when actually it is not (or even causes net harm). Whether beneficial or not, screening can create the false impression of an epidemic by increasing the number of cancers diagnosed. It can also produce a shift in the *proportion* of patients diagnosed at an early stage (even without a reduction in absolute incidence of late-stage disease) and inflate survival statistics without reducing mortality (i.e., the number of deaths from a given cancer relative to the number of those at risk for the cancer). In such a case, the *apparent* duration of survival (measured from date of diagnosis) increases without lives being saved or life expectancy changed.

Lead-time bias occurs whether or not a test influences the natural history of the disease; the patient is merely diagnosed at an earlier date. Survival *appears* increased even if life is not prolonged. The screening test only prolongs the time the subject is aware of the disease and spends as a cancer patient.

Length-biased sampling occurs because screening tests generally can more easily detect slow-growing, less aggressive cancers than fast-growing cancers. Cancers diagnosed due to the onset of symptoms between scheduled screenings are on average more aggressive, and treatment outcomes are not as favorable. An extreme form of length bias sampling is termed *overdiagnosis*, the detection of “pseudo disease.” The reservoir of some undetected slow-growing tumors is large. Many of these tumors fulfill the histologic criteria of cancer but will never become clinically significant or cause death during the patient’s remaining life span. This problem is compounded by the fact that the

most common cancers appear most frequently at ages when competing causes of death are more frequent.

Selection bias occurs because the population most likely to seek screening often differs from the general population to which the screening test might be applied. In general, volunteers for studies are more health conscious and likely to have a better prognosis or lower mortality rate, irrespective of the screening result. This is termed the *healthy volunteer effect*.

Potential Drawbacks of Screening Risks associated with screening include harm caused by the screening intervention itself, harm due to the further investigation of persons with positive tests (both true and false positives), and harm from the treatment of persons with a true-positive result, whether or not life is extended by treatment (e.g., even if a screening test reduces relative cause-specific mortality by 15–30%, 70–85% of those diagnosed still go on to die of the target cancer). The diagnosis and treatment of cancers that would never have caused medical problems can lead to the harm of unnecessary treatment and give patients the anxiety of a cancer diagnosis. The psychosocial impact of cancer screening can be substantial when applied to the entire population.

Assessment of Screening Tests Good clinical trial design can offset some biases of screening and demonstrate the relative risks and benefits of a screening test. A randomized controlled screening trial with cause-specific mortality as the endpoint provides the strongest support for a screening intervention. Overall mortality should also be reported to detect an adverse effect of screening and treatment on other disease outcomes (e.g., cardiovascular disease, treatment-induced cancers). In a randomized trial, two like populations are randomly established. One is given the usual standard of care (which may be no screening at all) and the other receives the screening intervention being assessed. Efficacy for the population studied is established when the group receiving the screening test has a better cause-specific mortality rate than the control group. Studies showing a reduction in the incidence of advanced-stage disease, improved survival, or a stage shift are weaker (and possibly misleading) evidence of benefit. These latter criteria are early indicators but not sufficient to establish the value of a screening test.

Although a randomized, controlled screening trial provides the strongest evidence to support a screening test, it is not perfect. Unless the trial is population-based, it does not remove the question of generalizability to the target population. Screening trials generally involve thousands of persons and last for years. Less definitive study designs are therefore often used to estimate the effectiveness of screening practices. However, every nonrandomized study design is subject to strong confounders. In descending order of strength, evidence may also be derived from the findings of internally controlled trials using intervention allocation methods other than randomization (e.g., allocation by birth date, date of clinic visit); the findings of analytic observational studies; or the results of multiple time series studies with or without the intervention.

Screening for Specific Cancers Screening for cervical, colon, and breast cancer has the potential to be beneficial for certain age groups. Depending on age and smoking history, lung cancer screening can also be beneficial in specific settings. Special surveillance of those at high risk for a specific cancer because of a family history or a genetic risk factor may be prudent, but few studies have assessed the effect on mortality. A number of organizations have considered whether or not to endorse routine use of certain screening tests. Because criteria have varied, they have arrived at different recommendations. The American Cancer Society (ACS) and the U.S. Preventive Services Task Force (USPSTF) publish screening guidelines (**Table 70-3**); the American Academy of Family Practitioners (AAFP) often follows/endorses the USPSTF recommendations; and the American College of Physicians (ACP) develops recommendations based on structured reviews of other organizations’ guidelines.

TABLE 70-3 Screening Recommendations for Asymptomatic Subjects Not Known to Be at Increased Risk for the Target Condition^a

CANCER TYPE	TEST OR PROCEDURE	USPSTF	ACS
Breast	Self-examination	"D" ^b (Not in current recommendations; from 2009)	Women, all ages: No specific recommendation
	Clinical examination	Women ≥40 years: "I" (as a stand-alone without mammography) (Not in current recommendations; from 2009)	Women, all ages: Do not recommend
	Mammography	Women 40–49 years: The decision to start screening mammography in women prior to age 50 years should be an individual one. Women who place a higher value on the potential benefit than the potential harms may choose to begin biennial screening between the ages of 40 and 49 years. ("C") Women 50–74 years: Every 2 years ("B") Women ≥75 years: "I"	Women 40–44 years: Provide the opportunity to begin annual screening Women 45–54 years: Screen annually Women ≥55 years: Transition to biennial screening or have the opportunity to continue annual screening Women ≥40 should continue screening mammography as long as their overall health is good and they have a life expectancy of 10 years or longer
	Magnetic resonance imaging (MRI)	"I" (Not in current recommendations; from 2009)	Women with >20% lifetime risk of breast cancer: Screen with MRI plus mammography annually Women with 15–20% lifetime risk of breast cancer: Discuss option of MRI plus mammography annually Women with <15% lifetime risk of breast cancer: Do not screen annually with MRI
	Tomosynthesis	Women, all ages: "I"	No specific recommendation
Cervical	Pap test (cytology)	Women <21 years: "D" Women 21–29 years: Screen with cytology alone every 3 years ("A") Women 30–65 years: Screen with cytology alone every 3 years, or with co-testing (HPV testing + cytology) every 5 years (two of three options, see HPV test below) ("A") Women >65 years, with adequate, normal prior Pap screenings: "D" Women after total hysterectomy for noncancerous causes: "D"	Women <21 years: No screening Women 21–29 years: Screen every 3 years Women 30–65 years: Screen with co-testing (HPV testing + cytology) every 5 years or cytology alone every 3 years (see HPV test below) Women >65 years: No screening following adequate negative prior screening Women after total hysterectomy for noncancerous causes: Do not screen
	HPV test	Women <30 years: Do not use HPV testing for cervical cancer screening Women 30–65 years: Screen with HPV testing alone or in combination with cytology every 5 years (two of three options, see Pap test above) ("A") Women >65 years, with adequate, normal prior Pap screenings: "D" Women after total hysterectomy for noncancerous causes: "D"	Women <30 years: Do not use HPV testing for cervical cancer screening Women 30–65 years: Preferred approach to screen with HPV and cytology co-testing every 5 years (see Pap test above) Women >65 years: No screening following adequate negative prior screening Women after total hysterectomy for noncancerous causes: Do not screen
Colorectal	Overall	Adults 50–75 years: "A" Screen for colorectal cancer; the risks and benefits of the different screening methods vary Adults 76–85 years: "C" The decision to screen should be an individual one, taking into account the patient's overall health and prior screening history	Adults ≥45–75 years: Screen for colorectal cancer with either a high-sensitivity stool-based test or a structural (visual) examination (≥45 years, qualified recommendation; ≥50 years, strong recommendation). Adults 76–85 years: Individualize screening based on patient preferences, life expectancy, health status, and prior screening history (qualified recommendation). Adults >85 years: Discourage screening (qualified recommendation). Every 5 years
	Sigmoidoscopy	Every 5 years; modeling suggests improved benefit if performed every 10 years in combination with annual FIT	Adults ≥45 years: Every 5 years
	Fecal occult blood testing (FOBT)	Every year	Adults ≥45 years: Every year
	Colonoscopy	Every 10 years	Adults ≥45 years: Every 10 years
	Fecal DNA testing	At least every 3 years	Adults ≥45 years: Every 3 years
	Fecal immunochemical testing (FIT)	Every year	Adults ≥45 years: Every year
	Computed tomography (CT) colonography	Every 5 years	Adults ≥45 years: Every 5 years

(Continued)

TABLE 70-3 Screening Recommendations for Asymptomatic Subjects Not Known to Be at Increased Risk for the Target Condition^a (Continued)

CANCER TYPE	TEST OR PROCEDURE	USPSTF	ACS
Lung	Low-dose CT scan	Adults 55–80 years, with a ≥30 pack-year smoking history, still smoking or have quit within past 15 years: "B" Discontinue once a person has not smoked for 15 years or develops a health problem that substantially limits life expectancy or the ability to have curative lung surgery	Men and women, 55–74 years, with ≥30 pack-year smoking history, still smoking or have quit within past 15 years: Discuss benefits, limitations, and potential harms of screening; offer smoking cessation counseling where relevant; only perform screening in high-volume, high-quality lung cancer screening and treatment centers.
Ovarian	CA-125 Transvaginal ultrasound	Women, all ages: "D" Women with a high-risk hereditary cancer syndrome: No recommendation	Currently, there are no reliable screening tests for the early detection of ovarian cancer. For women at high risk of ovarian cancer, it has not been proven that using transvaginal ultrasound or serum CA-125 lowers their chances of dying from ovarian cancer.
Prostate	Prostate-specific antigen (PSA)	Men 55–69 years: The decision to undergo periodic PSA-based screening should be an individual one. Men should have an opportunity to discuss the potential benefits and harms of screening with their clinician. Clinicians should not screen men who do not express a preference for screening ("C") Men ≥70 years: "D"	Starting at age 50, men at average risk and with a life expectancy of ≥10 years should talk to a doctor about the uncertainties, risks, and potential benefits of screening. If African American or have a father or brother who had prostate cancer before age 65, men should have this talk starting at age 45. For men with more than one first-degree relative with prostate cancer diagnosed before age 65, have this talk starting at age 40. How often they are screened will depend on their PSA level.
	Digital rectal examination (DRE)	No individual recommendation	As for PSA; if men decide to be tested, they should have the PSA blood test with or without a rectal exam.
Skin	Complete skin examination by clinician or patient	Adults, all ages: "I"	No guidelines

^aSummary of the screening procedures recommended for the general population by the USPSTF and the ACS. These recommendations refer to asymptomatic persons who are not known to have risk factors, other than age or gender, for the targeted condition. ^bUSPSTF lettered recommendations are defined as follows: "A": The USPSTF recommends the service because there is high certainty that the net benefit is substantial; "B": The USPSTF recommends the service because there is high certainty that the net benefit is moderate or moderate certainty that the net benefit is moderate to substantial; "C": The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences; there is at least moderate certainty that the net benefit is small; "D": The USPSTF recommends against the service because there is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits; "I": The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service.

Abbreviations: ACS, American Cancer Society; USPSTF, U.S. Preventive Services Task Force.

BREAST CANCER Breast self-examination, clinical breast examination by a caregiver, mammography, and magnetic resonance imaging (MRI) have all been variably advocated as useful screening tools.

A number of trials have suggested that annual or biennial screening with mammography in normal-risk women older than age 50 years decreases breast cancer mortality. Each trial has been criticized for design flaws. In most trials, breast cancer-related mortality rates were decreased by 15–30%. Experts disagree on whether average-risk women age 40–49 years should receive regular screening (Table 70-3). The U.K. Age Trial, the only randomized trial of breast cancer screening to specifically evaluate the impact of mammography in women age 40–49 years, found no statistically significant difference in breast cancer mortality for screened women versus controls after about 11 years of follow-up (relative risk 0.83; 95% confidence interval 0.66–1.04); however, <70% of women received screening in the intervention arm, potentially diluting the observed effect. A meta-analysis of nine large randomized trials showed an 8% relative reduction in mortality (relative risk 0.92; 95% confidence interval 0.75–1.02) from mammography screening for women age 39–49 years after 11–20 years of follow-up. This is equivalent to 3 breast cancer deaths prevented per 10,000 women >10 years (although the result is not statistically significant). At the same time, nearly half of women age 40–49 years screened annually will have false-positive mammograms necessitating further evaluation, often including biopsy. Estimates of overdiagnosis range from 10 to 50% of diagnosed invasive cancers. In the United States, widespread screening over the past several decades has not been accompanied by a reduction in incidence of metastatic breast cancer despite a large increase in early-stage disease, suggesting a substantial amount of overdiagnosis at the population level. In addition, the substantial improvements in systemic therapy have likely decreased the impact of mammography and early detection on falling breast cancer mortality rates.

Digital breast tomosynthesis is a newer method of breast cancer screening that reconstructs multiple x-ray images of the breast into superimposed "three-dimensional" slices. Although some evidence is available concerning the test characteristics of this modality, there are currently no data on its effects on health outcomes such as breast cancer-related morbidity, mortality, or overdiagnosis rates. A large randomized trial comparing standard digital mammography to tomosynthesis is in progress.

No study of breast self-examination has shown it to decrease mortality. A randomized controlled trial of approximately 266,000 women in China demonstrated no difference in breast cancer mortality between a group that received intensive breast self-exam instruction and reinforcement/reminders and controls at 10 years of follow-up. However, more benign breast lesions were discovered and more breast biopsies were performed in the self-examination arm.

Genetic screening for *BRCA1* and *BRCA2* mutations and other markers of breast cancer risk has identified a group of women at high risk for breast cancer. Unfortunately, when to begin and the optimal frequency of screening have not been defined. Mammography is less sensitive at detecting breast cancers in women carrying *BRCA1* and *BRCA2* mutations, possibly because such cancers occur in younger women, in whom mammography is known to be less sensitive. MRI screening may be more sensitive than mammography in women at high risk due to genetic predisposition or in women with very dense breast tissue, but specificity may be lower. An increase in overdiagnosis may accompany the higher sensitivity. The impact of MRI on breast cancer mortality with or without concomitant use of mammography has not been evaluated in a randomized controlled trial.

CERVICAL CANCER Screening with Papanicolaou (Pap) smears decreases cervical cancer mortality. The cervical cancer mortality rate has fallen substantially since the widespread use of the Pap smear. With

the onset of sexual activity comes the risk of sexual transmission of HPV, the fundamental etiologic factor for cervical cancer. Screening guidelines recommend regular Pap testing for all women who have reached the age of 21 (before this age, even in individuals that have begun sexual activity, screening may cause more harm than benefit). The recommended interval for Pap screening is 3 years. In all cases, screening more frequently adds little benefit but leads to important harms, including unnecessary procedures and overtreatment of transient lesions. Beginning at age 30, guidelines also include HPV testing with or without Pap smear. The screening interval for women who test normal using this approach may be lengthened to 5 years.

An upper age limit at which screening ceases to be effective is not known, but women age 65 years with no abnormal results in the previous 10 years may choose to stop screening. Screening should be discontinued in women who have undergone a hysterectomy with cervical excision for noncancerous reasons.

Although the efficacy of the Pap smear in reducing cervical cancer mortality has never been directly confirmed in a randomized, controlled setting, a clustered randomized trial in India evaluated the impact of one-time cervical visual inspection and immediate colposcopy, biopsy, and/or cryotherapy (where indicated) versus counseling on cervical cancer deaths in women age 30–59 years. After 7 years of follow-up, the age-standardized rate of death due to cervical cancer was 39.6 per 100,000 person-years in the intervention group versus 56.7 per 100,000 person-years in controls.

COLORECTAL CANCER Fecal occult blood testing (FOBT), digital rectal examination (DRE), rigid and flexible sigmoidoscopy, colonoscopy, and computed tomography (CT) colonography have been considered for colorectal cancer screening. A meta-analysis of five randomized controlled trials demonstrated a 22% relative reduction in colorectal cancer mortality after two to nine rounds of biennial FOBT at 30 years of follow-up; annual screening was shown to result in a greater colorectal cancer mortality reduction in a single trial (a 32% relative reduction). However, only 2–10% of those with occult blood in the stool have cancer. The high false-positive rate of FOBT substantially increases the number of colonoscopies performed.

Fecal immunochemical tests (FITs) have higher sensitivity for colorectal cancer than FOBT tests. Limited observational evidence suggests FITs may have lower ability to detect proximal versus distal colonic tumors. Multitargeted stool DNA testing combines FIT with testing for altered DNA biomarkers in cells that are shed into the stool. Although limited evidence demonstrates that it can have a higher single-test sensitivity for colorectal cancer than FIT alone, its specificity is lower, resulting in a higher number of false-positive tests and follow-up colonoscopies. There are no studies evaluating its effects on colorectal cancer incidence, morbidity, or mortality.

A blood test for the methylated *SEPT9* gene associated with colorectal cancer is available. However, its sensitivity is low, no longitudinal data have been collected on its performance or efficacy, and it is not recommended as a first-line screening test.

Two meta-analyses of five randomized controlled trials of sigmoidoscopy found an 18% relative reduction in colorectal cancer incidence and a 28% relative reduction in colorectal cancer mortality. Participant ages ranged from 50 to 74 years, with follow-up ranging from 6 to 13 years. Diagnosis of adenomatous polyps by sigmoidoscopy should lead to evaluation of the entire colon with colonoscopy. The most efficient interval for screening sigmoidoscopy is unknown, but an interval of 5 years is often recommended. Case-control studies suggest that intervals of up to 15 years may confer benefit; the randomized U.K. trial demonstrated colorectal cancer mortality reduction with one-time screening.

One-time colonoscopy detects ~25% more advanced lesions (polyps >10 mm, villous adenomas, adenomatous polyps with high-grade dysplasia, invasive cancer) than one-time FOBT with sigmoidoscopy; comparative *programmatic* performance of the two modalities over time is not known. Perforation rates are about 4/10,000 for colonoscopy and 1/10,000 for sigmoidoscopy. Debate continues on whether colonoscopy is too expensive and invasive and whether sufficient provider capacity exists to be recommended as the preferred screening

tool in standard-risk populations. Some observational studies suggest that efficacy of colonoscopy to decrease colorectal cancer mortality is primarily limited to the left side of the colon.

CT colonography, if done at expert centers, appears to have a sensitivity for polyps ≥6 mm, comparable to colonoscopy. However, the rate of extracolonic findings of abnormalities of uncertain significance that must nevertheless be worked up is high (~5–37%); the long-term cumulative radiation risk of repeated colonography screenings is also a concern.

LUNG CANCER Chest x-ray and sputum cytology have been evaluated in several randomized lung cancer screening trials. The most recent and largest ($n = 154,901$) of these, a component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial, found that, compared with usual care, annual chest x-ray did not reduce the risk of dying from lung cancer (relative risk 0.99; 95% confidence interval 0.87–1.22) after 13 years. However, it showed evidence of overdiagnosis associated with chest x-ray. Low-dose CT has also been evaluated in several randomized trials. The largest and longest of these, the National Lung Screening Trial (NLST), was a randomized controlled trial of screening for lung cancer in ~53,000 persons age 55–74 years with a 30+ pack-year smoking history. It demonstrated a statistically significant reduction of about 3 fewer deaths per 1000 people screened with CT compared to chest x-ray after 12 years. However, the harms include the potential radiation risks associated with multiple scans, the discovery of incidental findings of unclear significance, and a high rate of false-positive test results. Both incidental findings and false-positive tests can lead to invasive diagnostic procedures associated with anxiety, expense, and complications (e.g., pneumo- or hemothorax after lung biopsy). The NLST was performed at experienced screening centers, and the balance of benefits and harms may differ in the community setting at less experienced centers.

OVARIAN CANCER Adnexal palpation, transvaginal ultrasound (TVUS), and serum CA-125 assay have been considered for ovarian cancer screening. A large randomized, controlled trial has shown that an annual screening program of TVUS and CA-125 in average-risk women does not reduce deaths from ovarian cancer (relative risk 1.21; 95% confidence interval 0.99–1.48). Adnexal palpation was dropped early in the study because it did not detect any ovarian cancers that were not detected by either TVUS or CA-125. A second large randomized trial used a two-stage screening approach incorporating a risk of ovarian cancer algorithm that determined whether additional testing with CA-125 or TVUS was required. At 14 years of follow-up, there was no statistically significant reduction in ovarian cancer deaths. The risks and costs associated with the high number of false-positive results are impediments to routine use of these modalities for screening. In the PLCO trial, 10% of participants had a false-positive result from TVUS or CA-125, and one-third of these women underwent a major surgical procedure; the ratio of surgeries to screen-detected ovarian cancer was approximately 20:1. In September 2016, the FDA issued a safety communication recommending against using any screening test, including the risk of ovarian cancer algorithm, for ovarian cancer.

PROSTATE CANCER The most common prostate cancer screening modalities are digital rectal exam (DRE) and serum PSA assay. An emphasis on PSA screening has caused prostate cancer to become the most common nonskin cancer diagnosed in American males. This disease is prone to lead-time bias, length bias, and overdiagnosis, and substantial debate continues among experts as to whether screening should be offered unless the patient specifically asks to be screened. Virtually all organizations stress the importance of informing men about the uncertainty regarding screening efficacy and the associated harms. Prostate cancer screening clearly detects many asymptomatic cancers, but the ability to distinguish tumors that are lethal but still curable from those that pose little or no threat to health is limited, and randomized trials indicate that the effect of PSA screening on prostate cancer mortality across a population is, at best, small. Men older than age 50 years have a high prevalence of indolent, clinically insignificant prostate cancers (about 30–50% of men, increasing further as men age).

Two major randomized controlled trials of the impact of PSA screening on prostate cancer mortality have been published. The PLCO Cancer Screening Trial was a multicenter U.S. trial that randomized almost 77,000 men age 55–74 years to receive either annual PSA testing for 6 years or usual care. At 13 years of follow-up, no statistically significant difference in the number of prostate cancer deaths was noted between the arms (rate ratio 1.09; 95% confidence interval 0.87–1.36). More than half of men in the control arm received at least one PSA test during the trial, which may have diluted a small effect.

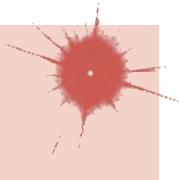
The European Randomized Study of Screening for Prostate Cancer (ERSPC) was a multinational study that randomized ~182,000 men between age 50 and 74 years (with a predefined “core” screening group of men age 55–69 years) to receive PSA testing or no screening. Recruitment and randomization procedures, as well as actual frequency of PSA testing, varied by country. After a median follow-up of 15.5 years, a 20% relative reduction in the risk of prostate cancer death in the screened arm was noted in the “core” screening group. The trial found that 570 men (95% confidence interval 380–1137 men) would need to be invited to screening, and 18 cases of prostate cancer detected, to avert 1 death from prostate cancer. There was an unexplained imbalance in treatment between the two study arms, with a higher proportion of men with clinically localized cancer receiving radical prostatectomy in the screening arm and receiving it at experienced referral centers.

Screening must be linked to effective therapy in order to have any benefit. Two trials conducted after the initiation of widespread PSA testing did not find a substantial decrease in prostate cancer deaths in control arms of “watchful waiting” or monitoring (i.e., no curative treatment) compared to radical prostatectomy or radiation therapy. Prostate cancer-specific survival was very good (about 99%) and nearly identical at a median follow-up of 10 years. Treatments for low-stage prostate cancer, such as surgery and radiation therapy, can cause substantial morbidity, including impotence and urinary incontinence.

SKIN CANCER Visual examination of all skin surfaces by the patient or by a health care provider is used in screening for basal and squamous cell cancers and melanoma. No prospective randomized study has been performed to look for a mortality decrease. Unfortunately, screening is associated with a substantial rate of overdiagnosis.

FURTHER READING

- FENTON JJ et al: Prostate-specific antigen-based screening for prostate cancer: Evidence report and systematic review for the U.S. Preventive Services Task Force. *JAMA* 319:1914, 2018.
- KRAMER BS, CROSWELL JM: Cancer screening: The clash of science and intuition. *Annu Rev Med* 60:125, 2009.
- MANSON JE et al: Vitamin D supplements and prevention of cancer and cardiovascular disease. *N Engl J Med* 380:33, 2019.
- MCNEIL JJ et al: Effect of aspirin on all-cause mortality in the healthy elderly. *N Engl J Med* 379:1519, 2018.
- MELNIKOW J et al: Screening for cervical cancer with high-risk human papillomavirus testing: Updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA* 320:687, 2018.
- NATIONAL LUNG SCREENING TRIAL RESEARCH TEAM: Lung cancer incidence and mortality with extended follow-up in the National Lung Screening Trial. *J Thorac Oncol* 14:1732, 2019.
- NELSON HD: Effectiveness of breast cancer screening: Systematic review and meta-analysis to update the 2009 U.S. Preventive Services Task Force recommendation. *Ann Intern Med* 164:244, 2016.
- US PREVENTIVE SERVICES TASK FORCE et al: Screening for colorectal cancer: US Preventive Services Task Force recommendation. *JAMA* 325:1965, 2021.
- WELCH HG et al: Epidemiologic signatures in cancer. *N Engl J Med* 384:14, 2019.
- ZERAATKAR D et al: Effect of lower versus high red meat intake on cardiometabolic and cancer outcomes: A systematic review of randomized trials. *Ann Intern Med* 171:721, 2019.



CANCER IS A GENETIC DISEASE

Cancer arises through a series of somatic alterations in DNA that result in unrestrained cellular proliferation. Most of these alterations involve subtle sequence changes in DNA (i.e., mutations). The somatic mutations may originate as a consequence of random replication errors or exposure to carcinogens (e.g., radiation) and can be exacerbated by faulty DNA repair processes. While most cancers arise sporadically, clustering of cancers occurs in families that carry a germline mutation in a cancer gene.

HISTORICAL PERSPECTIVE

The idea that cancer progression is driven by sequential somatic mutations in specific genes has only gained general acceptance in the past 30 years. Before the advent of the microscope, cancer was believed to be composed of aggregates of mucus or other noncellular matter. By the middle of the nineteenth century, it became clear that tumors were masses of cells and that these cells arose from the normal cells of the tissue from which the cancer originated. The molecular basis for the uncontrolled proliferation of cancer cells was to remain a mystery for another century. During that time, a number of theories for the origin of cancer were postulated. The great biochemist Otto Warburg proposed the combustion theory of cancer, which stipulated that cancer was due to abnormal oxygen metabolism. Others believed that all cancers were caused by viruses and that cancer was in fact a contagious disease.

In the end, observations of cancer occurring in chimney sweeps, studies of x-rays, and the overwhelming data demonstrating cigarette smoke as a causative agent in lung cancer, together with Ames's work on chemical mutagenesis, were consistent with the idea that cancer originated through changes in DNA. However, it was not until the somatic mutations responsible for cancer were identified at the molecular level that the genetic basis of cancer was definitively established. Although the viral theory of cancer did not prove to be generally accurate (with exceptions such as human papillomaviruses, which can cause cervical and other cancers), the study of retroviruses led to the discovery of the first human *oncogenes* in the late 1970s. Oncogenes are one of the two major classes of cancer driver genes. The study of families with genetic predisposition to cancer was instrumental to the discovery of the other major class of cancer driver genes, called *tumor-suppressor genes*. Current technologies permit the sequence analysis of entire cancer genomes and provide a comprehensive view of the genetic changes that cause tumors to arise and become malignant. The field that studies the various types of mutations, as well as the consequences of these mutations in tumor cells, is now known as *cancer genetics*.

THE CLONAL ORIGIN AND MULTISTEP NATURE OF CANCER

Nearly all cancers originate from a single cell; this clonal origin is a critical discriminating feature between neoplasia and hyperplasia. Multiple cumulative mutational events are invariably required for the progression of a tumor from normal to fully malignant phenotype. The process can be seen as Darwinian microevolution in which, at each successive step, the mutated cells gain a growth advantage resulting in the expansion of a neoplastic clone (Fig. 71-1). Based on observations of cancer frequency increases during aging, the epidemiologists Armitage and Doll and Nordling independently proposed that cancer is a result of three discrete cellular changes. Remarkably, this early model has been validated by extensive sequencing of cancer genomes. These studies revealed that just three causal mutations are required for the development of several of the most common cancers. Overall, it is currently believed that most common solid tumors require a minimum of three mutated cancer driver genes (either oncogenes or tumor-suppressor genes) for their development. One or two mutations are

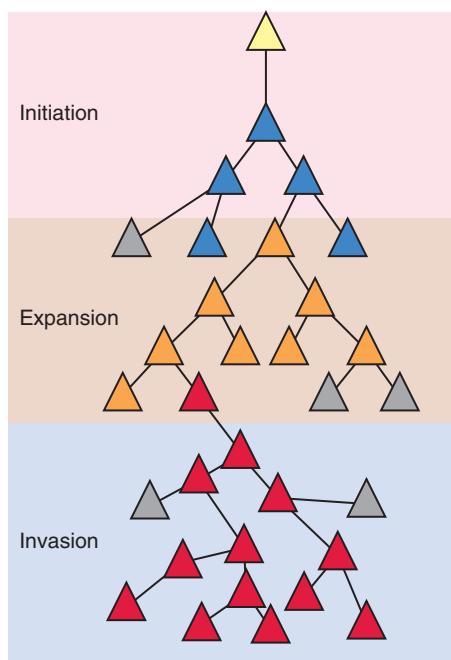


FIGURE 71-1 Multistep clonal development of malignancy. In this diagram, a series of three cumulative mutations, each with a modest growth advantage acting alone, eventually results in a malignant tumor. Note that not all such alterations result in progression. The actual number of cumulative mutations necessary to transform from the normal to the malignant state has been estimated to be three for several of the most common types of cancer. (Adapted and modified from PC Nowell: *The clonal evolution of tumor cell populations*. *Science* 194:23, 1976.)

sufficient for benign tumorigenesis, but not for the invasive capacity that distinguishes cancers from benign tumors. Less common tumors, such as liquid tumors (leukemias or lymphomas), sarcomas, and childhood tumors, appear to require only two driver gene alterations for malignancy. Note that a cancer driver gene is best defined as one containing a mutation that increases the selective growth advantage of the cell containing it. Normally, cell birth and cell death are in perfect equilibrium; every time a cell is born, another in the same lineage dies. Cancer driver gene mutations alter this equilibrium, so that more cells are born than die. The imbalance is often slight, so that the difference between cell birth and cell death can be less than 1%. This explains, in combination with the low rate of mutation, why tumorigenesis—the journey from a normal cell to a typical malignant, solid tumor—often takes decades.

We now know the precise nature of the genetic alterations responsible for nearly all malignancies and are beginning to understand how these alterations promote the distinct stages of tumor growth. The prototypical example is colon cancer, in which analyses of genomes from the entire spectrum of neoplastic growths—from normal colon epithelium through adenoma to carcinoma—have identified mutations that are highly characteristic of each type of lesion (Fig. 71-2).

TWO TYPES OF CANCER GENES: ONCOGENES AND TUMOR-SUPPRESSOR GENES

Oncogenes and tumor-suppressor genes exert their effects on tumor growth through their ability to determine cell fates, influence cell survival, and contribute to genome maintenance. The underlying molecular mechanisms can be extremely complex. While tightly regulated in normal cells, oncogenes acquire mutations that typically relieve this control and lead to increased activity of the

gene products. This activating mutational event occurs in a single allele. In contrast, the normal function of tumor-suppressor genes is usually to restrain cell growth, and this function is lost in cancer. Because of the diploid nature of mammalian cells, both alleles must be inactivated for a cell to completely lose the function of a tumor-suppressor gene. Thus, it requires two genetic events to inactivate a tumor-suppressor gene mutation, while only one genetic event is required to activate an oncogene.

A subset of tumor-suppressor genes controls the ability of the cell to maintain the integrity of its genome. Cells with a deficiency in these genes acquire an increased number of mutations throughout their genomes, including those in oncogenes and tumor-suppressor genes. This “mutator” phenotype was first hypothesized by Loeb to explain how the multiple rare mutational events required for tumorigenesis can occur in the lifetime of an individual. A mutator phenotype underlies several forms of cancer, such as those associated with deficiencies in DNA mismatch repair. The great majority of cancers do not harbor repair deficiencies, and their rate of mutation is similar to that observed in normal cells. Many of these cancers, however, appear to harbor a different kind of genetic instability, affecting the loss or gains of whole chromosomes or large parts thereof (as explained in more detail below).

ONCOGENES IN HUMAN CANCER

Work by Peyton Rous in the early 1900s revealed that a chicken sarcoma could be transmitted from animal to animal in cell-free extracts, suggesting that cancer could be induced by an agent acting positively to promote tumor formation. The agent responsible for the transmission of the cancer was a retrovirus (Rous sarcoma virus [RSV]), and the oncogene responsible was identified 75 years later as V-SRC. Other oncogenes were also discovered through their presence in the genomes of retroviruses that are capable of causing cancers in chickens, mice, and rats. The nonmutated cellular homologues of these viral genes are called proto-oncogenes and are often targets of mutation or aberrant regulation in human cancer. Whereas many oncogenes were discovered on the basis of their presence in retroviruses, other oncogenes, particularly those involved in translocations characteristic of particular leukemias and lymphomas, were identified through genomic approaches. Investigators cloned the sequences surrounding the chromosomal translocations observed cytogenetically and identified the genes activated at the breakpoints (see below). Some of these were oncogenes previously found in retroviruses (like *ABL*, involved in chronic myeloid leukemia [CML]), whereas others were new (like *BCL2*, involved in B-cell lymphoma). In the normal cellular environment, proto-oncogenes have crucial roles in cell proliferation and

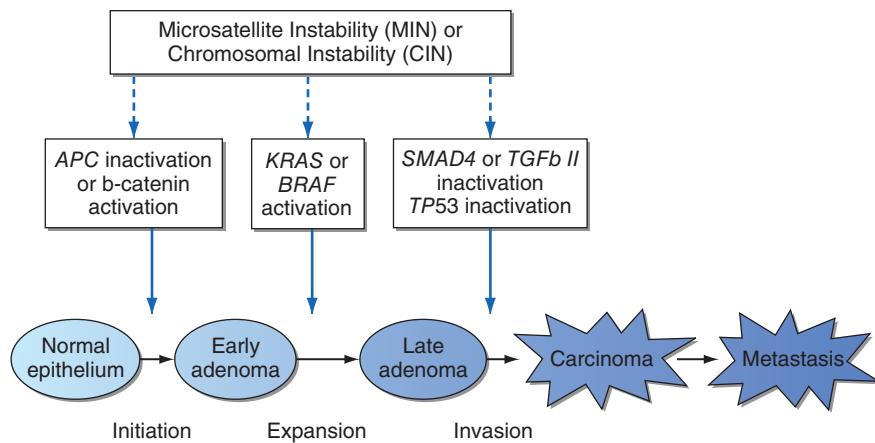


FIGURE 71-2 Progressive somatic mutational steps in the development of colon carcinoma. The accumulation of alterations in a number of different genes results in the progression from normal epithelium through adenoma to full-blown carcinoma. Genetic instability (microsatellite or chromosomal) accelerates the progression by increasing the likelihood of mutation at each step. Patients with familial polyposis are already one step into this pathway, because they inherit a germline alteration of the *APC* gene. TGF, transforming growth factor.

TABLE 71-1 Oncogenes Commonly Altered in Human Cancers

ONCOGENE	FUNCTION	ALTERATION IN CANCER	NEOPLASM
<i>AKT1</i>	Serine/threonine kinase	Point mutation	Skin
<i>BRAF</i>	Serine/threonine kinase	Point mutation	Melanoma, thyroid, colorectal
<i>CCND1</i>	Cell cycle progression	Amplification	Esophageal, head and neck
<i>CTNNB1</i>	Signal transduction	Point mutation	Colon, liver, uterine, melanoma
<i>EGFR</i>	Signal transduction	Point mutation	Lung
<i>FLT3</i>	Signal transduction	Point mutation	AML
<i>IDH1</i>	Chromatin modification	Point mutation	Glioma
<i>MDM2</i>	Inhibitor of p53	Amplification	Sarcoma, glioma
<i>MDM4</i>	Inhibitor of p53	Amplification	Breast
<i>MYC</i>	Transcription factor	Amplification	Prostate, ovarian, breast, liver, pancreatic
<i>MYCL1</i>	Transcription factor	Amplification	Ovarian, bladder
<i>MYCN</i>	Transcription factor	Amplification	Neuroblastoma
<i>PIK3CA</i>	Phosphoinositol-3-kinase	Point mutation	Multiple cancers
<i>KRAS</i>	GTPase	Point mutation	Pancreatic, colorectal, lung
<i>NRAS</i>	GTPase	Point mutation	Melanoma

Abbreviation: AML, acute myeloid leukemia.

differentiation. **Table 71-1** is a partial list of oncogenes known to be involved in human cancer.

The normal growth and differentiation of cells is controlled by growth factors that bind to receptors on the surface of the cell. The signals generated by the membrane receptors are transmitted inside the cells through signaling cascades involving kinases, G proteins, and other regulatory proteins. Ultimately, these signals affect the activity of transcription factors in the nucleus, which regulate the expression of genes crucial in cell proliferation, cell differentiation, and cell death. Oncogene products function at critical steps in these signaling pathways (**Chap. 72**). Inappropriate activation of these pathways can lead to tumorigenesis.

MECHANISMS OF ONCOGENE ACTIVATION

POINT MUTATION

Point mutation (alternatively known as single nucleotide substitution) is a common mechanism of oncogene activation. For example, mutations in *KRAS* are present in >95% of pancreatic cancers and 40% of colon cancers. Activating *KRAS* mutations are less common in other cancer types, although they can occur at significant frequencies in leukemia, lung, and thyroid cancers. Remarkably—and in contrast to the diversity of mutations found in tumor-suppressor genes—most of the activated *KRAS* alleles contain point mutations in codons 12, 13, or 61. These mutations lead to constitutive activation of the mutant RAS protein. The restricted pattern of mutations observed in oncogenes compared to that of tumor-suppressor genes reflects the fact that gain-of-function mutations must occur at specific sites, while a broad variety of mutations can lead to loss of activity. Indeed, inactivation of a gene can in theory be accomplished through the introduction of a stop codon anywhere in the coding sequence, whereas activations require precise substitutions at residues that can somehow lead to an increase in the activity of the encoded protein under particular circumstances within the cell.

DNA AMPLIFICATION

The second mechanism for activation of oncogenes is DNA sequence amplification, leading to overexpression of the gene product. This increase in DNA copy number may cause cytologically recognizable chromosome alterations referred to as *homogeneous staining regions* (HSRs) if integrated within chromosomes, or *double minutes* (dmins) if extrachromosomal. With microarray and DNA sequencing technologies, the entire genome can be surveyed for gains and losses of DNA sequences, thus pinpointing chromosomal regions likely to contain genes important in the development or progression of cancer.

Numerous genes have been reported to be amplified in cancer. Several of these genes, including *NMYC* and *LMYC*, were identified

through their presence within the amplified DNA sequences of a tumor and their homology to known oncogenes. Because amplified regions often include hundreds of thousands of base pairs, multiple oncogenes may be amplified in a single amplicon in some cancers. For example, *MDM2*, *GLI1*, *CDK4*, and *TPSPAN31* at chromosomal location 12q13-15 have been shown to be co-amplified in several types of sarcomas and other tumors; which of these genes play the causal role in the neoplastic process is still an active area of research. Amplification of a cellular gene is often a predictor of poor prognosis; for example, *ERBB2/HER2* and *NMYC* are often amplified in aggressive breast cancers and neuroblastoma, respectively.

CHROMOSOMAL REARRANGEMENT

Chromosomal alterations provide important clues to the genetic changes in cancer. The chromosomal alterations in human solid tumors such as carcinomas are heterogeneous

and complex and occur as a result of the frequent chromosomal instability observed in these tumors (see below). In contrast, the chromosome alterations in myeloid and lymphoid tumors are often simple translocations, that is, reciprocal transfers of chromosome arms from one chromosome to another. The breakpoints of recurring chromosome abnormalities usually occur at the site of cellular oncogenes. **Table 71-2** lists representative examples of recurring chromosome alterations in malignancy and the associated gene(s) rearranged or deregulated by the chromosomal rearrangement. Translocations are often observed in liquid tumors in general and are particularly common in lymphoid tumors, probably because these cell types have the capability to rearrange their DNA to generate antigen receptors. Indeed, antigen receptor genes are commonly involved in the translocations, implying that an imperfect regulation of receptor gene rearrangement may be involved in their pathogenesis. In addition to transcription factors and signal transduction molecules, translocation may result in the overexpression of cell cycle regulatory proteins or proteins such as cyclins and of proteins that regulate cell death. Recurrent translocations have more recently been identified in solid tumors such as prostate cancers. For example, fusions between *TMPRSS2* and *ERG*, which are normally located in tandem on chromosome 21, contribute to more than one-third of all prostate cancers.

The first reproducible chromosome abnormality detected in human malignancy was the Philadelphia chromosome detected in CML.

TABLE 71-2 Representative Oncogenes at Chromosomal Translocations

GENE (CHROMOSOME)	TRANSLOCATION	MALIGNANCY
<i>BCR-ABL</i>	(9;22)(q34;q11)	Chronic myeloid leukemia
<i>BCL1(11q13.3)-IgH</i> (14q32)	(11;14)(q13;q32)	Mantle cell lymphoma
<i>BCL2(18q21.3)-IgH</i> (14q32)	(14;18)(q32;q21)	Follicular lymphoma
<i>FLI-EWSR1</i>	(11;22)(q24;q12)	Ewing's sarcoma
<i>LCK-TCRB</i>	(1;7)(p34;q35)	T-cell acute lymphocytic leukemia
<i>PAX3-FOX01</i>	(2;13)(q35;q14)	Rhabdomyosarcoma
<i>PAX8-PPARG</i>	(2;3)(q13;p25)	Thyroid
<i>IL21R-BCL6</i>	(3;16)(q27;p11)	Non-Hodgkin's lymphoma
<i>TAL1-TCTA</i>	(1;3)(p34;p21)	Acute T-cell leukemia
<i>TMPRSS2-ERG</i>	Rearrangement on Chr21q22	Prostate

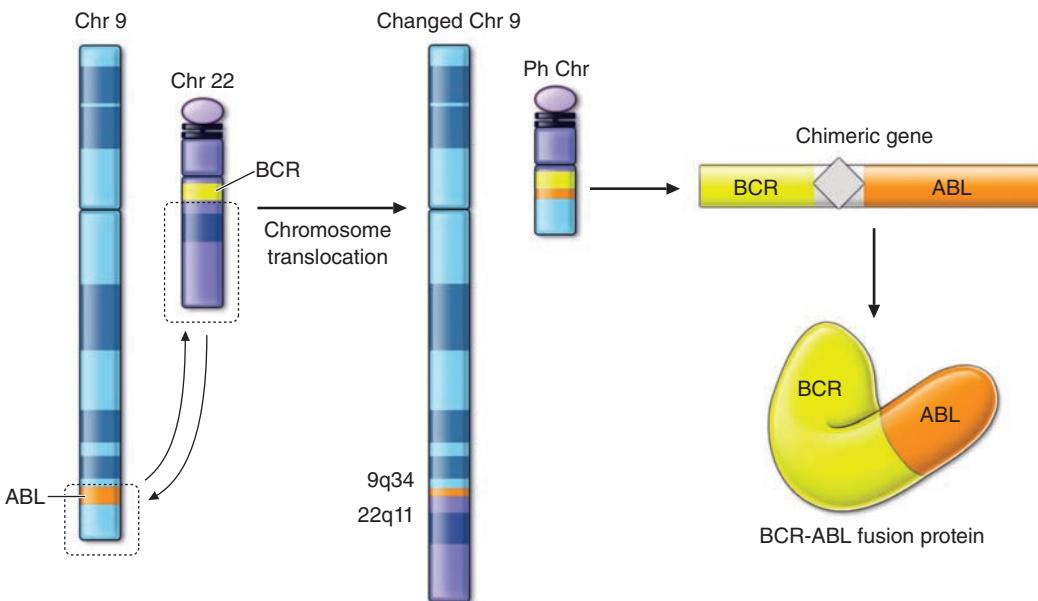


FIGURE 71-3 Specific translocation seen in chronic myeloid leukemia (CML). The Philadelphia chromosome (Ph) is derived from a reciprocal translocation between chromosomes 9 and 22 with the breakpoint joining the sequences of the *ABL* oncogene with the *BCR* gene. The fusion of these DNA sequences allows the generation of an entirely novel fusion protein with modified function.

This cytogenetic abnormality is generated by reciprocal translocation involving the *ABL* oncogene on chromosome 9, encoding a tyrosine kinase, being placed in proximity to the breakpoint cluster region (*BCR*) gene on chromosome 22. Figure 71-3 illustrates the generation of the translocation and its protein product. The consequence of expression of the *BCR-ABL* gene product is the activation of signal transduction pathways leading to cell growth independent of normal external signals. Imatinib (marketed as Gleevec), a drug that specifically blocks the activity of Abl tyrosine kinase, has shown remarkable efficacy with little toxicity in patients with CML. The successful targeting of *BCR-ABL* by imatinib is the paradigm for molecularly targeted anticancer therapies.

CHROMOSOMAL INSTABILITY IN SOLID TUMORS

Solid tumors generally contain an abnormal number of chromosomes, a state known as aneuploidy; chromosomes from aneuploid tumors exhibit structural alterations such as translocations, deletions, and amplifications. These abnormalities reflect an underlying defect in cancer cells known as chromosomal instability. While aneuploidy is a striking cellular phenotype, chromosomal instability is manifest as only a small increase in the tendency of cells to gain, lose, or rearrange chromosomes during any given cell cycle. This intrinsically low rate of chromosome aberration implies that cancer cells become aneuploid only after many generations of clonal expansion. The molecular basis of aneuploidy remains incompletely understood. It is widely believed that defects in checkpoints, the quality-control mechanisms that halt the cell cycle if chromosomes are damaged or misaligned, contribute to chromosomal instability. This hypothesis emerged from experimental observations that the tumor suppressor p53 controls checkpoints that regulate the initiation of DNA replication and the onset of mitosis. These processes are therefore defective in many cancer cells. The mitotic spindle checkpoint, which ensures proper chromosome attachment to the mitotic spindle before allowing the sister chromatids to separate, is also altered in some cancers, irrespective of p53 status. The precise relationship between checkpoint deficiency, p53, and chromosomal instability remains unclear, but it is believed that even a subtle perturbation of the highly orchestrated process of cell division can impact the ability of a cell to faithfully replicate and segregate its complement of chromosomes. From a therapeutic standpoint, the checkpoint defects that are prevalent in cancers have been proposed as

vulnerabilities that may be exploited by novel agents and combinatorial strategies.

In contrast to the genome-wide cytogenetic changes that are typical indications of an underlying chromosomal instability, more focal patterns of chromosomal rearrangement have been recurrently detected in several cancer types. A curious phenomenon known as *chromothripsis* causes dozens of distinct breakpoints that are localized on one or several chromosomes. These striking structural alterations are thought to reflect a single event in which a chromosome is fragmented and then imprecisely reassembled. While the exact process that underlies chromothripsis remains obscure and its effects on driver genes are not yet clear, a transient period of extreme instability stands in contrast to the gradual loss, gain, and rearrangement of chromosomes that are typically observed in serially cultured cancer cells.

TUMOR-SUPPRESSOR GENE INACTIVATION IN CANCER

The first functional evidence for tumor-suppressor genes came from experiments showing that fusion of mouse cancer cells with normal mouse fibroblasts led to a nontumorigenic phenotype in the fused cells. The normal role of tumor-suppressor genes is to restrain cell growth, and the function of these genes is inactivated in cancer. The three major types of somatic lesions observed in tumor-suppressor genes during tumor development are *point mutations*, small insertions and/or deletions known as *indels*, and *large deletions*. Point mutations or indels in the coding region of tumor-suppressor genes will frequently lead to truncated protein products or allele-specific loss of RNA expression by the process of *nonsense-mediated decay*. Unlike the highly recurrent point mutations that are found in critical positions of activated oncogenes, known as mutational *hotspots*, the point mutations that cause tumor-suppressor gene inactivation tend to be distributed throughout the open reading frame. Large deletions lead to the loss of a functional product and sometimes encompass the entire gene or even the entire chromosome arm, leading to loss of heterozygosity (LOH) in the tumor DNA compared to the corresponding normal tissue DNA (Fig. 71-4). LOH in tumor DNA often indicates the presence of a tumor-suppressor gene at a particular chromosomal location, and LOH studies have been useful in the positional cloning of many tumor-suppressor genes. The rate of LOH is increased in the presence of chromosomal instability, a relationship that would explain the selective forces leading to the high prevalence of aneuploidy in late-stage cancers.

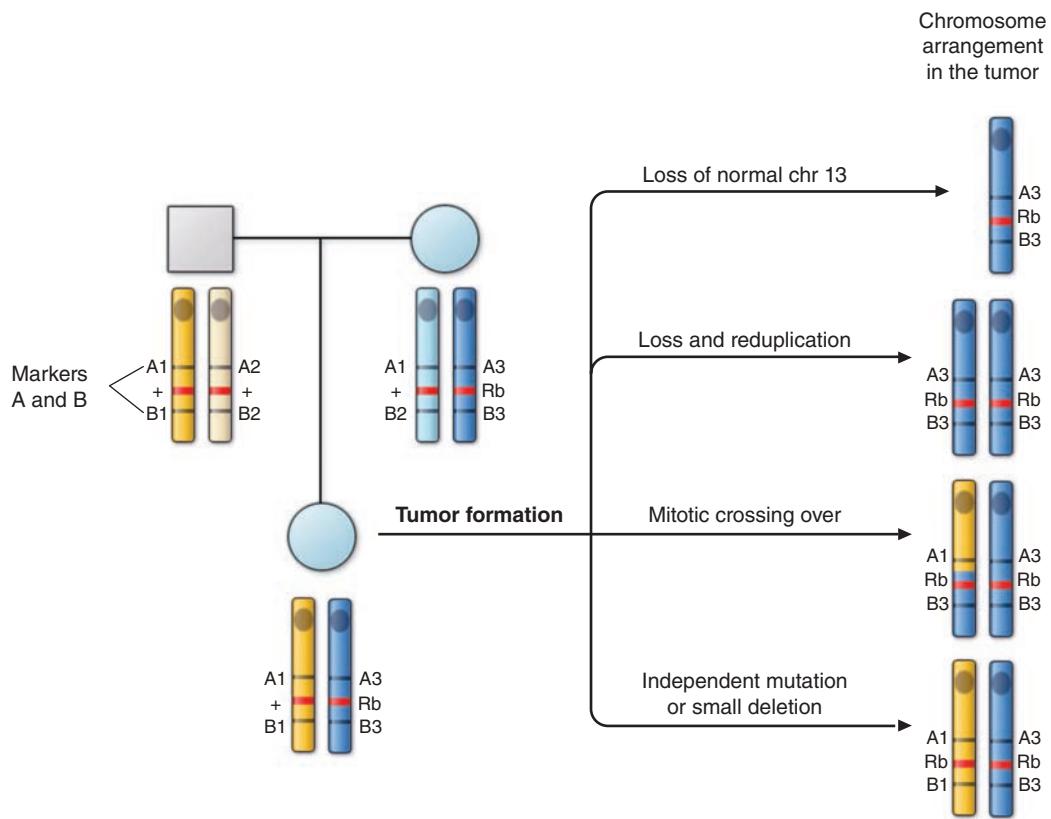


FIGURE 71-4 Diagram of possible mechanisms for tumor formation in an individual with hereditary (familial) retinoblastoma. On the left is shown the pedigree of an affected individual who has inherited the abnormal (Rb) allele from her affected mother. The normal allele is shown as a (+). The four chromosomes of her two parents are drawn to indicate their origin. Flanking the retinoblastoma locus are genetic markers (A and B) also analyzed in this family. Markers A3 and B3 are on the chromosome carrying the retinoblastoma disease gene. Tumor formation results when the normal allele, which this patient inherited from her father, is inactivated. On the right are shown four possible ways in which this could occur. In each case, the resulting chromosome 13 arrangement is shown. Note that in the first three situations, the normal allele (B1) has been lost in the tumor tissue, which is referred to as loss of heterozygosity (LOH) at this locus.

Gene silencing, an epigenetic change that leads to the loss of gene expression, occurs in conjunction with hypermethylation of the promoter and histone deacetylation, and is another mechanism of tumor-suppressor gene inactivation. An *epigenetic modification* refers to a covalent modification of chromatin, heritable by cell progeny that may involve DNA but does not involve a change in the DNA sequence. The inactivation of the second X chromosome in female cells is a physiologic example of an epigenetic silencing that prevents gene expression from the inactivated chromosome. Genomic regions of hypermethylated and hypomethylated DNA can be detected by specialized techniques, and a subset of these regional modifications has consequences on the cell's behavior.

FAMILIAL CANCER SYNDROMES

A small fraction of cancers occurs in patients with a genetic predisposition. Based on studies of inherited and sporadic forms of retinoblastoma, Knudson and others formulated a hypothesis that explains the differences between sporadic and inherited forms of the same tumor type. In inherited forms of cancer, called *cancer predisposition syndromes*, one allele of a particular tumor-suppressor gene is inherited in mutant form. This germline mutation is not sufficient to initiate a tumor, however; the other allele, inherited from the unaffected parent, must become somatically inactivated in a normal stem cell for tumorigenesis to be initiated. In sporadic (noninherited) forms of the same disease, all cells in the body start out with two normal copies of the tumor-suppressor gene. A single cell must then sequentially acquire mutations in both alleles of the tumor-suppressor gene to initiate a tumor. Thus, biallelic mutations of the same tumor-suppressor gene are required for both inherited and noninherited forms of the disease; the only difference is that individuals with the inherited form have a "head start": they already have one allele mutated, from conception, and only need one additional mutation to initiate the process (Fig. 71-4).

This distinction explains why those with inherited forms of the disease develop more cancers, at an earlier age, than the general population. It also explains why, even though every cell in an individual with a cancer predisposition syndrome has a mutant gene, only a relatively small number of tumors arise during his or her lifetime. The reason is that the vast majority of cells within such individuals are functionally normal because one of the two alleles of the tumor-suppressor gene is normal. Mutations are uncommon events, and only the rare cells that develop a mutation in the remaining normal allele will exhibit uncontrolled proliferation. The same principle applies to virtually all types of cancer predisposition syndromes, though the particular genes differ. For example, inherited mutations in *RB1*, *WT1*, *VHL*, *APC*, and *BRCA1* lead to predispositions to retinoblastomas, Wilms' tumors, renal cell carcinomas, colorectal carcinomas, and breast carcinomas, respectively (Table 71-3). Also note that the biallelic inactivation of any of these genes is not sufficient to develop cancer; it requires other, additional somatic alterations in other genes for the initiating cells to evolve to malignancy, as noted above.

Roughly 100 familial cancer syndromes have been reported; the great majority are very rare. Most of these syndromes exhibit an autosomal dominant pattern of inheritance, although some of those associated with DNA repair abnormalities (xeroderma pigmentosum, Fanconi's anemia, ataxia telangiectasia) are inherited in an autosomal recessive fashion. Table 71-3 shows a number of cancer predisposition syndromes and the responsible genes.

The next section examines inherited colon cancer predispositions in detail because several lessons of general importance have been derived from the study of these syndromes.

Familial adenomatous polyposis (FAP) is a dominantly inherited colon cancer syndrome caused by germline mutations in the adenomatous polyposis coli (*APC*) tumor-suppressor gene on chromosome 5. Affected individuals develop hundreds to thousands of adenomas in

TABLE 71-3 Cancer Predisposition Syndromes and Associated Genes

SYNDROME	GENE	CHROMOSOME	INHERITANCE	TUMORS
Ataxia telangiectasia	ATM	11q22-q23	AR	Breast
Autoimmune lymphoproliferative syndrome	FAS	10q24 1q23	AD	Lymphomas
	FASL			
Bloom's syndrome	BLM	15q26.1	AR	Various
Cowden's syndrome	PTEN	10q23	AD	Breast, thyroid
Familial adenomatous polyposis	APC	5q21	AD	Colorectal (early onset)
	MUTYH	1p34.1	AR	
Familial melanoma	CDKN2A	9p21	AD	Melanoma, pancreatic
Familial Wilms' tumor	WT1	11p13	AD	Kidney (pediatric)
Hereditary breast/ovarian cancer	BRCA1	17q21	AD	Breast, ovarian, prostate
	BRCA2	13q12.3		
Hereditary diffuse gastric cancer	CDH1	16q22	AD	Stomach
Hereditary multiple exostoses	EXT1	8q24	AD	Exostoses, chondrosarcoma
	EXT2	11p11-12		
Hereditary retinoblastoma	RB1	13q14.2	AD	Retinoblastoma, osteosarcoma
Hereditary nonpolyposis colon cancer (HNPCC)	MSH2	2p16	AD	Colon, endometrial, ovarian, stomach, small bowel, ureter carcinoma
	MLH1	3p21.3		
	MSH6	2p16		
	PMS2	7p22		
Hereditary papillary renal carcinoma	MET	7q31	AD	Papillary kidney
Juvenile polyposis syndrome	SMAD4	18q21	AD	Gastrointestinal, pancreatic
	BMPR1A			
Li-Fraumeni syndrome	TP53	17p13.1	AD	Sarcoma, breast
Multiple endocrine neoplasia type 1	MEN1	11q13	AD	Parathyroid, endocrine, pancreas, and pituitary
Multiple endocrine neoplasia type 2a	RET	10q11.2	AD	Medullary thyroid carcinoma, pheochromocytoma
Neurofibromatosis type 1	NF1	17q11.2	AD	Neurofibroma, neurofibrosarcoma, brain
Neurofibromatosis type 2	NF2	22q12.2	AD	Vestibular schwannoma, meningioma, spine
Nevoid basal cell carcinoma syndrome (Gorlin's syndrome)	PTCH1	9q22.3	AD	Basal cell carcinoma, medulloblastoma, jaw cysts
Tuberous sclerosis	TSC1	9q34	AD	Angiofibroma, renal angiomyolipoma
	TSC2	16p13.3		
von Hippel–Lindau disease	VHL	3p25-26	AD	Kidney, cerebellum, pheochromocytoma

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

the colon. In each of these adenomas, the *APC* allele inherited from the affected parent has been inactivated by virtue of a somatic mutation (Fig. 71-2). This inactivation usually occurs through a gross chromosomal event resulting in loss of all or a large part of the long arm of chromosome 5, where *APC* resides. In other cases, the remaining allele is inactivated by a subtle intragenic mutation of *APC*, which is typically a single base substitution resulting in a nonsense codon. Gross chromosomal losses occur more commonly than point mutations in normal cells, explaining why chromosomal loss rather than point mutation is the predominant mechanism underlying the inactivation of the normal allele of *APC*. The same is true for other cancer predisposition syndromes caused by other inherited tumor suppressor gene mutations; gross chromosomal events are generally responsible for inactivation of the tumor-suppressor gene allele inherited from the nonaffected parent. Several thousand adenomas form in FAP patients, and a small subset of the millions of cells within an adenoma will acquire a second mutation, leading to tumor progression, that is, a larger adenoma. A third mutation in such a larger adenoma may convert it to a carcinoma. If untreated (by colectomy), at least one of the adenomas will progress to cancer by the time patients are in their mid-40s. *APC* can be considered to be a gatekeeper for colon tumorigenesis in that in the absence of mutation of this gatekeeper (or a gene acting within the same pathway), a colorectal tumor simply cannot be initiated. Figure 71-5 shows the germline and somatic mutations found in the *APC* gene. A negative regulator of a signaling pathway that determines cell fate during development, the *APC* protein provides differentiation and apoptotic cues

to colonic epithelial cells as they migrate up the crypt. Defects in this process can lead to abnormal accumulation of cells that would otherwise differentiate and eventually undergo apoptosis.

In contrast to patients with FAP, patients with hereditary nonpolyposis colon cancer (HNPCC, or Lynch's syndrome) do not develop polyposis, but instead develop only one or a small number of adenomas that rapidly progress to cancer. HNPCC is due to inherited mutations in one of four DNA mismatch repair genes (Table 71-3) that are components of a repair system responsible for correcting errors in newly replicated DNA. Germline mutations in *MSH2* and *MLH1* account for more than 90% of HNPCC cases, and mutations in *MSH6* and *PMS2* account for the remainder. When a somatic mutation inactivates the remaining wild-type allele of a mismatch repair gene, the cell develops a hypermutable phenotype characterized by profound genomic instability that is most readily apparent in short repeated sequences called *microsatellites* and is sometimes called microsatellite instability (MSI). The high rate of mutation in such cells impacts all genes, including oncogenes and tumor suppressor genes, and thereby accelerates the activation of the former and the inactivation of the latter (Fig. 71-2). HNPCC can be considered a disease of tumor progression; once tumors are initiated (by an inactivating mutation of *APC* or by some other gene in the *APC* pathway), tumors rapidly progress because of the accelerated mutation rate. Progression from a tiny adenoma to carcinoma takes only a few years in HNPCC patients instead of the two or three decades this progression takes in patients with FAP (or in patients with sporadic colorectal tumors). Approximately half of

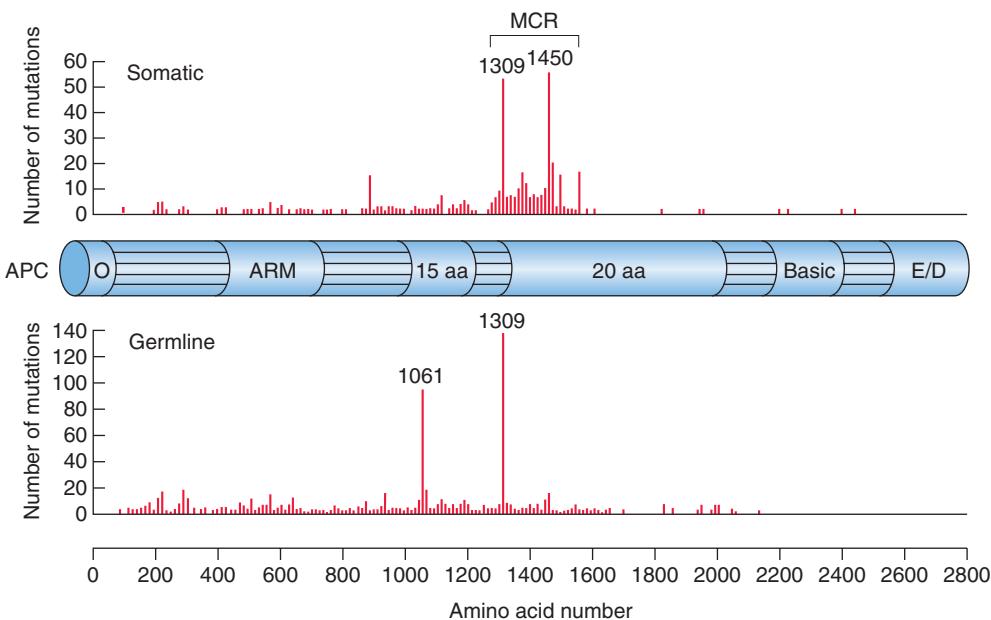


FIGURE 71-5 Germline and somatic mutations in the tumor-suppressor gene adenomatous polyposis coli (APC). APC encodes a 2843-amino-acid protein with six major domains: an oligomerization region (O), armadillo repeats (ARM), 15-amino-acid repeats (15 aa), 20-amino-acid repeats (20 aa), a basic region, and a domain involved in binding EB1 and the *Drosophila* discs large homologue (E/D). Shown are 650 somatic and 826 germline mutations representative of the mutations that occur within the APC gene (from the APC database at www.umd.be/APC). All known pathogenic mutations of APC result in the truncation of the APC protein. Germline mutations are found to be relatively evenly distributed up to codon 1600 except for two mutation hotspots surrounding amino acids 1061 and 1309, which together account for one-third of the mutations found in familial adenomatous polyposis (FAP) families.

HNPCC patients develop colorectal cancers by the time they are in their mid-40s—similar to that of FAP patients. This coincidence in age of onset emphasizes that both tumor initiation (abnormal in FAP patients) and tumor progression (abnormal in HNPCC patients) are the two pillars of cancer development and are equally important for cancer development.

Another general principle is apparent from the comparison between FAP and HNPCC patients. The tumors in FAP patients, like those in patients without hereditary predisposition to cancers, exhibit chromosomal instability rather than MSI. Indeed, MSI and chromosomal instability tend to be mutually exclusive in colon cancers, suggesting that they represent alternative mechanisms for the generation of genomic instability (Fig. 71-2). Other cancer types rarely exhibit MSI. Chromosomal instability is far more prevalent than MSI among all cancer types, perhaps explaining why nearly all cancers are aneuploid.

Although most autosomal dominant inherited cancer syndromes are due to mutations in tumor-suppressor genes (Table 71-3), there are a few interesting exceptions. Multiple endocrine neoplasia type 2, a dominant disorder characterized by pituitary adenomas, medullary carcinoma of the thyroid, and (in some pedigrees) pheochromocytoma, is due to gain-of-function mutations in the proto-oncogene *RET* on chromosome 10. Similarly, gain-of-function mutations in the tyrosine kinase domain of the *MET* oncogene lead to hereditary papillary renal carcinoma. Interestingly, loss-of-function mutations in the *RET* gene cause a completely different disease, Hirschsprung's disease (aganglionic megacolon [Chaps. 328 and 388]).

Although the heritable forms of cancer have taught us much about the mechanisms of growth control, most forms of cancer do not follow simple Mendelian patterns of inheritance. The majority of human cancers arise in a sporadic fashion, solely as a result of somatic mutation, and in the absence of any mutations in cancer-predisposing genes in their germlines.

GENETIC TESTING FOR FAMILIAL CANCER

The discovery of cancer susceptibility genes raises the possibility of DNA testing to predict the risk of cancer in individuals of affected families. An algorithm for cancer risk assessment and decision making

in high-risk families using genetic testing is shown in Fig. 71-6. Once a mutation is discovered in a family, subsequent testing of asymptomatic family members can be crucial in patient management. A negative gene test in these individuals can prevent years of anxiety, providing comfort in the knowledge that their cancer risk is no higher than that of the general population. On the other hand, a positive test may lead to alteration of clinical management, such as increased frequency of cancer screening and, when feasible and appropriate, prophylactic surgery. Potential negative consequences of a positive test result include psychological distress (anxiety, depression) and discrimination, although the Genetic Information Nondiscrimination Act (GINA) makes it illegal for predictive genetic information to be used to discriminate in health insurance or employment. Testing should therefore not be conducted without counseling before testing is administered and during and after disclosure of the test result.

It is now feasible to obtain high-quality sequence of all of the protein-coding DNA sequences, and even of the entire genome, in any given individual. In such studies, numerous variants in DNA sequences will inevitably be identified in every subject, but the significance of the vast majority of these DNA sequence findings will be unclear. Even mutations in tumor-suppressor genes will be difficult to interpret unless there is an obvious functional implication, such as the truncation of the open reading frame, or that particular mutation has previously been correlated with cancer in other individuals. Germline mutations associated with cancer predisposition are uncommon in individuals without a family history of cancer, though they do occur. Much more common are *variants of unknown significance* (VUS). VUS that are found during genetic testing cannot be used to evaluate the relative risk of cancer but may nonetheless cause anxiety because they represent a deviation from the reference allele that is established as “normal.” Because of the low yield of informative mutations that modify cancer risk and the frequent identification of VUS, it is generally not appropriate to use DNA sequencing to assess cancer risk in individuals without a family history of cancer. However, there are exceptions: *testing may be appropriate in some subpopulations with a known increased risk, even without a personal family history*. For example, two mutations in the breast cancer susceptibility gene *BRCA1*, 185delAG and 5382insC, exhibit a sufficiently high frequency in the Ashkenazi

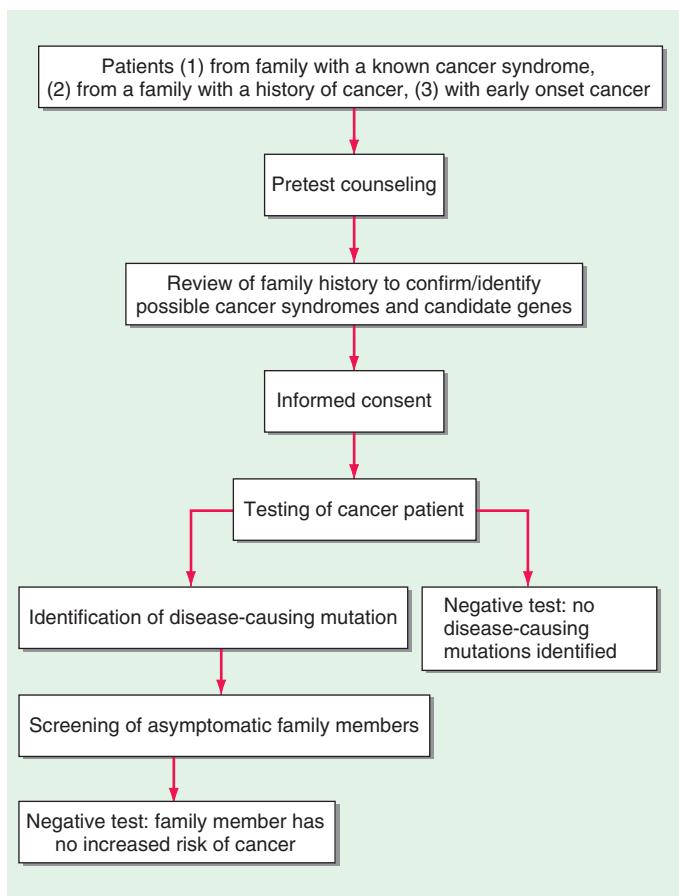


FIGURE 71-6 Algorithm for genetic testing in a family with cancer predisposition. The key step is the identification of a disease mutation in a cancer patient, which is an indication for the testing of asymptomatic family members. Asymptomatic family members who test positive may require increased screening or surgery, whereas those who test negative are at no greater risk for cancer than the general population. It should be emphasized that no molecular assay used for this sort of testing is 100% sensitive; negative results must be interpreted with this caveat in mind.

Jewish population that genetic testing based on ethnicity alone may be warranted.

It is important that genetic test results be communicated to families by trained genetic counselors. To ensure that the families clearly understand its advantages and disadvantages and the impact it may have on disease management and psyche, genetic testing should never be done before counseling. Significant expertise is needed to communicate the results of genetic testing to families.

VIRUSES IN HUMAN CANCER

Several human malignancies are associated with viruses. Examples include Burkitt's lymphoma (Epstein-Barr virus; [Chap. 194](#)), hepatocellular carcinoma (hepatitis viruses), cervical cancer (human papillomavirus [HPV]; [Chap. 198](#)), and T-cell leukemia (retroviruses; [Chap. 201](#)). There are several types of HPV, including the high-risk types 16 and 18 that are strongly associated with the development of cervical, vulvar, vaginal, penile, anal, and oropharyngeal cancer. The mechanisms of action of all these viruses involve inactivation of tumor-suppressor genes. For example, HPV proteins E6 and E7 bind to and inactivate cellular tumor suppressors p53 and pRB, respectively. This is the reason that HPV is such a potent initiator of cancer: infection with a virus is tantamount to having two of the three mutant driver genes required for cancer, that is, one viral oncogene inactivates p53 and the other inactivates Rb. Once these two inactivated gene products initiate tumorigenesis, only one additional mutant gene is required for these tumors to progress to malignancy.

CANCER GENOMES

The advent of relatively inexpensive technologies for rapid and high-throughput DNA sequencing has facilitated the comprehensive analysis of numerous genomes from many types of tumors. This unprecedented view into the genetic nature of cancer has provided remarkable insights. Most cancers do not arise in the context of a mutator phenotype, and accordingly, the number of mutations in even the most advanced cancers is relatively modest. Common solid tumors harbor 30–70 subtle mutations that are nonsynonymous (i.e., result in an amino acid change in the encoded protein). Liquid tumors such as leukemias, as well as pediatric tumors, typically have fewer than 20 mutations. The vast majority of the mutations detected in tumors are not functionally significant; they simply arose by chance in a single cell that gave rise to an expanding clone. Such mutations, which provide no selective advantage to the cell in which they occur, are known as *passenger* mutations. As noted above, only a small number of the mutations confer a selective growth advantage and thereby promote tumorigenesis. These functional mutations are known as *driver* mutations, and the genes in which they occur are called driver genes.

The frequency and distribution of driver mutations within a single tumor type can be represented as a topographical landscape. The picture that emerges from these studies reveals that most genes that are mutated in tumors are actually mutated at relatively low frequencies, as would be expected of passenger genes, whereas a small number of genes (the driver genes) are mutated in a large proportion of tumors. It appears that there are only a total of ~120 bona fide driver genes contributing to the development of solid tumors of all kinds, though other driver genes that play a role in a small fraction of cancers are still being discovered. The majority of the mutations in driver genes provide a direct selective growth advantage by altering the signaling pathways that mediate cell survival or the determination of cell fate. The remaining driver gene mutations indirectly provide a selective growth advantage by accelerating the mutation rate of proto-oncogenes and tumor-suppressor genes. That the same driver genes play a role in multiple cancer types was unexpected before their discovery and has important implications for the development of new “tumor-agnostic” therapeutic and diagnostic approaches. Moreover, the functions of all these driver genes can be organized into a small number of signaling pathways, as shown in [Table 71-4](#).

As a consequence of the mutations they harbor, cancer cells invariably express mutant proteins that are only rarely found in normal cells. Some of these mutant proteins are processed and displayed on the cell surface in the context of major histocompatibility complexes, a process that would normally facilitate their recognition by the adaptive immune system. Thus, all cancers have the theoretical potential to be recognized as foreign, or “nonself,” via the display of these tumor-specific antigens, known as mutation-associated neoantigens (MANAs). In fact, established tumors invariably prevent the activation of local

TABLE 71-4 Signaling Pathways Altered in Cancer

PROCESS	PATHWAY	REPRESENTATIVE DRIVER GENES
Cell survival	Cell cycle regulation/apoptosis	<i>RB1, BCL2</i>
	RAS	<i>KRAS, BRAF</i>
	PIK3CA	<i>PTEN, PIK3CA</i>
	JAK/STAT	<i>JAK2, FLT3</i>
	MAPK	<i>MAP3K, ERK</i>
	TGF-β	<i>BMPR1A, SMAD4</i>
Cell fate	Notch	<i>NOTCH1, FBXW7</i>
	Hedgehog	<i>PTCH1, SMO</i>
	WNT/APC	<i>APC, CTNNB1</i>
	Chromatin modification	<i>DNMT1, IDH1</i>
	Transcriptional regulation	<i>AR, KLF4</i>
Genome maintenance	DNA damage signaling and repair	<i>ATM, BRCA1</i>

T cells by inducing an intercellular suppressive mechanism known as an *immune checkpoint*. Therapeutic approaches to exploit this potential vulnerability by blocking immune checkpoints have elicited striking responses in patients with several types of cancer.

It was hypothesized that the potential immunogenicity of a tumor would be related to the total number of distinctive neoantigens it can express, which in turn is directly determined by the total number of mutations in the cancer genome. This does seem to be the case. Colorectal cancers that develop as a result of mismatch repair deficiency and smoking-related lung cancers, both of which characteristically harbor large numbers of mutations, exhibit more robust responses to therapeutic immune checkpoint blockade than most other tumor types. Notably, driver mutations as well as passenger mutations that result in the expression of mutant proteins can both contribute to the display of immunogenic neoantigens. Thus, the total number of coding mutations, a metric known as *mutational load*, is one of the determinants of potential immunogenicity.

TUMOR HETEROGENEITY

The mutant cells that compose a single tumor are not genetically identical. Rather, cells obtained from different sites on a tumor will harbor common mutations as well as mutations that are unique to each sample. Genetic heterogeneity results from the ongoing acquisition of mutations during tumor growth. Each time a genome is replicated, there is a small but quantifiable probability that a mutation will spontaneously arise as a result of a replication error and be passed on to the cellular progeny. This is true in normal cells or in tumor cells. Any randomly chosen cell from the skin of one individual will harbor hundreds of genetic alterations that distinguish it from a different randomly chosen skin cell, and the same is true for all organs of self-renewing tissues. Tumors are actually *less* genetically heterogeneous than normal tissues; any two randomly chosen cells from a tumor of an individual will have fewer differences than any two randomly chosen cells from that individual's normal tissues. The reason for this decrease in heterogeneity is

clonal expansion, the fundamental feature of tumorigenesis. Every time a clonal expansion occurs, a genetic bottleneck wipes out heterogeneity among the cells that did not expand; these unexpanded cells either die or form only a minute proportion of the total cells in the expanding tumor.

The mutations that vary between cells of a given tumor are invariably passenger mutations that arose since the last evolutionary bottleneck, that is, those mutations that arose during the expansion of the founder cell that gave rise to the final clonal expansion. In contrast, the passenger mutations that were present in the founder cell will be uniformly present in every cell in the tumor. In that respect, these passenger mutations are not heterogeneously distributed and are in fact uniformly present in virtually all cancer cells. These "clonal" mutations, i.e., present in all cells of the cancers, are the main source of MANAs that can be exploited through immune checkpoint inhibitors. The total number of mutations and their distribution within tumor cells represent a complex interplay between the age of the patient (the older the patient, the more passenger mutations will have accumulated in the founding cell of the first clonal expansion) and the evolutionary history of the cancer (its age and number of clonal expansions it experienced).

Tumor heterogeneity has been recognized for decades at the cytogenetic, biochemical, and histopathologic levels. However, it is only recently, with the advent of a deep understanding of cancer genetics, that genetic heterogeneity can be interpreted in a medically relevant fashion. The first important point to recognize about tumor heterogeneity is that it is only the variation in driver gene alterations that is important; the cellular distribution of passenger gene mutations is irrelevant except for immune-related phenomena. In this discussion of heterogeneity, we can expand the definition of "driver genes" to include those that provide a selective growth advantage in the face of therapy in addition to those that provide a selective growth advantage during tumor evolution, prior to treatment.

Type I heterogeneity refers to that among tumors of the same type from different patients (Fig. 71-7). Though adenocarcinomas of the

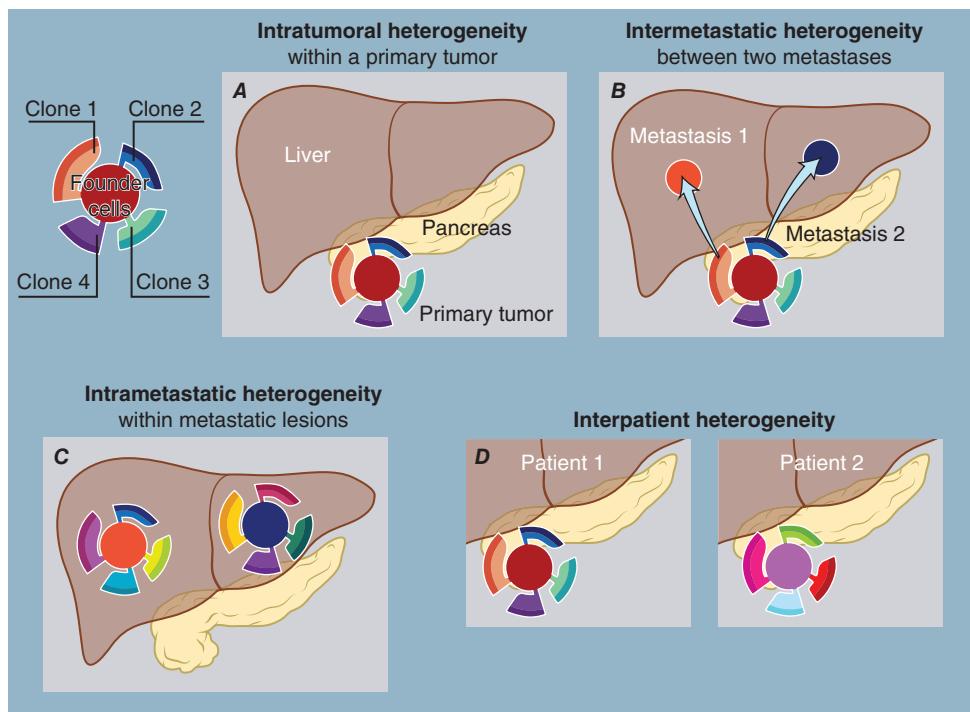


FIGURE 71-7 Four types of tumor heterogeneity. Tumor heterogeneity is the inevitable result of cell proliferation, as new mutations are introduced during clonal expansion. In a typical tumor (*upper left*), founder cells that harbor a large fraction of the total mutations give rise to subclones, which continue to evolve independently. The tumors of the founding populations are shown in the middle of each circle; the distinct subclones are shown around the periphery. **A.** Heterogeneity among the cells of a primary tumor is known as intratumoral heterogeneity. **B.** Heterogeneity among the founding cells of distinct metastatic lesions (marked as 1 and 2) that arise in the same patient is known as intermetastatic heterogeneity. **C.** Heterogeneity among the cells of each metastatic tumor is known as intrametastatic heterogeneity. **D.** Interpatient heterogeneity. The mutations in the tumors of two patients are almost completely distinct. (From B Vogelstein et al: Cancer genome landscapes. *Science* 339:1546, 2013. Reprinted with permission from AAAS.)

lung generally harbor mutations in three or more driver genes, the genes differ among the patients, and the precise mutations within the same gene can vary considerably. Type I heterogeneity is the basis for precision medicine, where the goal is to treat patients with drugs that target the proteins encoded by genetic alterations within their specific tumors. Type II heterogeneity refers to the genetic heterogeneity among different cells from the same primary tumor. Tumors continue to evolve as they grow, and different cells of the same cancer, in its original site (e.g., the pancreas), may acquire other driver gene mutations that are not shared among the other cells of the tumor. Such a mutation can result in a small clonal expansion that may or may not be important biologically. In cases in which the primary tumor can be surgically excised, such mutations are unimportant unless they give rise to type III heterogeneity (described below). The reason they are unimportant is because all primary tumor cells, whether homogeneous or not, are removed by the surgical procedure. In primary tumors that cannot be completely excised (such as most advanced brain tumors and many pancreatic ductal adenocarcinomas), heterogeneity is biomedically important because it can give rise to drug resistance, analogously to that described for type IV heterogeneity (see below). Type III heterogeneity refers to the genetic differences among the founder cells of the metastatic lesions from the same patient. For example, a patient with melanoma may have 100 different metastases distributed throughout various organs. Only if a mutant *BRAF* is present in every founder cell of every metastasis, then the patient has a chance at a complete response to a *BRAF* inhibitor. There have been several recent detailed studies of the metastases from various tumor types. Fortunately, these studies suggest there is very little, if any, type III heterogeneity among driver genes, a necessary prerequisite for the successful implementation of current and future targeted therapies. Finally, type IV heterogeneity refers to that among cells of individual metastatic lesions. As the founder cell of each metastasis expands to become detectable, it acquires mutations, a small number of which can act as “drivers” when the patient is exposed to therapeutics. This type of heterogeneity is of major clinical importance, as it has been shown to be responsible for the development of resistance in virtually all targeted therapies. The development of such resistance is a fait accompli based simply on known mutation rates and genetic resistance mechanisms. The only way to circumvent acquired resistance is to treat metastatic tumors earlier (i.e., in adjuvant setting, before much tumor expansion has occurred) or to treat with combinations of drugs for which cross-resistance is genetically impossible.

PERSONALIZED CANCER DETECTION AND TREATMENT

High-throughput DNA sequencing has led to an unprecedented understanding of cancer at the molecular level. A comprehensive mutation profile provides a molecular history of a given tumor and insights into how it arose. Because tumor cells and tumor DNA are shed into the blood and other bodily fluids, common driver mutations can be used as highly specific biomarkers for early detection. For diagnosed tumors, tumor-specific mutations can be used to estimate tumor burden, assess treatment responses, and detect recurrence.

In some cases, information regarding specific genes and pathways that are altered provides patients and physicians with options for personalized therapy. This general approach is sometimes referred to as *precision medicine*. Because tumor behavior is highly variable, even within a tumor type, personalized information-based medicine can supplement and perhaps eventually supplant histology-based tumor assessment, especially in the case of tumors that are resistant to conventional therapeutic approaches. Conversely, molecular nosology has revealed similarities in tumors of diverse histotype. The success

of the precision medicine approach in any given patient depends on the presence of tumor-associated genetic alterations that are actionable (i.e., can be targeted with a specific drug). Examples of currently actionable changes include mutations in *BRAF* (targeted by the drug vemurafenib), *RET* (targeted by sunitinib and sorafenib), *ALK* rearrangements (targeted by crizotinib), and mismatch repair genes (targetable by immune checkpoint inhibitors). At present, the proportion of tumors that can be treated with such precision medicine approaches is relatively small, but future therapeutic development will hopefully change this situation. The development of new targeted agents is at present hindered by the fact that most such agents can only target activated oncogenes, while the great majority of genetic alterations in common solid tumors are those that inactivate tumor-suppressor genes. Because all drugs, whether for use in oncology or any other purpose, can only inhibit protein actions, drugs cannot be used to directly target the proteins encoded by inactivated tumor-suppressor genes; these proteins are already inactive. More information about the pathways through which tumor-suppressor genes act may provide a way around this obstacle. For example, when a tumor-suppressor gene is inactivated, some downstream component of the pathway is likely to be activated, thereby presenting a realistic target. An example of this is provided by PARP-1 inhibitors, which have been successfully used to treat patients whose tumors have inactivating mutations of genes involved in DNA repair processes, such as *BRCA1*. Patterns of global gene expression can be used to help unravel such pathways and are already being used to predict drug sensitivities and provide prognostic information in addition to that provided by DNA sequence analysis. Evaluation of proteomic and metabolomics patterns may also prove useful for this purpose.

THE FUTURE

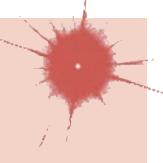
A revolution in cancer genetics has occurred in the past 30 years. Most types of cancer are now understood at the DNA sequence level, and this accomplishment has led to an increasingly refined understanding of tumorigenesis. Cancer gene mutations have proven to be reliable biomarkers for cancer detection and monitoring as well as for informing therapeutics through precision medicine approaches. Gene-based tests are already standard of care for patients with certain tumor types, such as melanoma and colorectal and lung cancers, and the utility of these tests will undoubtedly be expanded in the coming years as new therapies and ways of predicting responses to therapies are developed. While effective treatment of advanced cancers remains difficult, the early promise shown by immune-based therapies notwithstanding, it is expected that breakthroughs in these areas will continue to emerge and be applicable to an ever-increasing number of cancers. Moreover, with the hoped-for advances in diagnostics, particularly in the earlier detection of cancers, the new and old therapies for cancer can be expected to have a much greater impact on reducing cancer deaths.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the past contributions of Pat J. Morin, Jeff Trent, and Francis Collins to earlier versions of this chapter.

FURTHER READING

- BUNZ F: *Principles of Cancer Genetics*, 2nd ed. Dordrecht, Springer, 2016.
- LE DT et al: PD-1 Blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 372:2509, 2015.
- VOGELSTEIN B et al: Cancer genome landscapes. *Science* 339:1546, 2013.
- VOGELSTEIN B, KINZLER KW: The path to cancer—three strikes and you’re out. *N Engl J Med* 373:1895, 2015.

**TABLE 72-1 Phenotypic Characteristics of Malignant Cells**

Deregulated cell proliferation: Loss of function of negative growth regulators (tumor suppressor genes, i.e., *Rb*, *p53*), and increased action of positive growth regulators (oncogenes, i.e., *Ras*, *Myc*). Leads to aberrant cell cycle control and includes loss of normal checkpoint responses.

Failure to differentiate: Arrest at a stage before terminal differentiation. May retain stem cell properties. (Frequently observed in leukemias due to transcriptional repression of developmental programs by the gene products of chromosomal translocations.)

Loss of normal apoptosis pathways: Inactivation of *p53*, increases in *Bcl-2* (antiapoptotic) family members. This defect enhances the survival of cells with oncogenic mutations and genetic instability and allows clonal expansion and diversification within the tumor without activation of physiologic cell death pathways.

Genetic instability: Defects in DNA repair pathways leading to either single nucleotide or oligonucleotide mutations (as in microsatellite instability, MIN) or, more commonly, chromosomal instability (CIN) leading to aneuploidy (abnormal number of chromosomes in a cell). Caused by loss of function of a number of proteins including *p53*, *BRCA1/2*, mismatch repair genes, DNA repair enzymes, and the spindle checkpoint. Leads to accumulation of a variety of mutations in different cells within the tumor and heterogeneity.

Loss of replicative senescence: Normal cells stop dividing in vitro after 25–50 population doublings. Arrest is mediated by the *Rb*, *p16^{INK4a}*, and *p53* pathways. While most cells remain arrested, genetic and epigenetic changes in a subset of cells allow further replication, leading to telomere loss, with crisis leading to death of many cells. Cells that survive often harbor gross chromosomal abnormalities and the ability to continue to proliferate. These cells express telomerase, which maintains telomeres and is important for ongoing growth of these cells. Relevance to human *in vivo* cancer remains uncertain. Many human cancers express telomerase.

Nonresponsiveness to external growth-inhibiting signals: Cancer cells have lost responsiveness to signals normally present to stop proliferating when they have overgrown the niche normally occupied by the organ from which they are derived. Our understanding about this mechanism of growth regulation remains limited.

Increased angiogenesis: Due to increased gene expression of proangiogenic factors (VEGF, FGF, IL-8, angiopoietin) by tumor or stromal cells, or loss of negative regulators (endostatin, tumstatin, thrombospondin).

Invasion: Cell mobility and ability to move through extracellular matrix and into other tissues or organs. Loss of cell-cell contacts (gap junctions, cadherins) and increased production of matrix metalloproteinases (MMPs). Can take the form of epithelial-to-mesenchymal transition (EMT), with anchored epithelial cells becoming more like motile fibroblasts.

Metastasis: Spread of tumor cells to lymph nodes or distant tissue sites. Limited by the ability of tumor cells to migrate out of initial site and to survive in a foreign environment, including evading the immune system (see below).

Evasion of the immune system: Downregulation of MHC class I and II molecules; induction of T-cell tolerance; inhibition of normal dendritic cell and/or T-cell function; antigenic loss variants and clonal heterogeneity; increase in regulatory T cells.

Shift in cell metabolism: Complex changes including alterations due to tumor stress such as hypoxia and energy generation shifts from oxidative phosphorylation to aerobic glycolysis generate building blocks for malignant cell production and proliferation.

Abbreviations: FGF, fibroblast growth factor; IL, interleukin; MHC, major histocompatibility complex; VEGF, vascular endothelial growth factor.

normal. Cancer cells clearly have lost responsiveness to such controls and do not recognize when they have overgrown the niche normally occupied by the organ from which they are derived. A better understanding of these mechanisms of growth regulation in the context of organ homeostasis is evolving.

CANCER AS AN ORGAN THAT IGNORES ITS NICHE

The fundamental cellular defects that create a malignant neoplasm act at the cellular level, and some of these are cell autonomous. However, that is not the entire story. Cancers consist of both malignant cells as well as other cells, blood vessels, extracellular matrix, and signaling and other molecules in the cancer microenvironment. They behave as organs that have lost their specialized function and stopped responding

CANCER CELL BIOLOGY

Cancers are characterized by unregulated cell division, avoidance of cell death, tissue invasion, and the ability to spread to other areas of the body (metastasize). A neoplasm is *benign* when it grows in an unregulated fashion without tissue invasion or metastasizing. The presence of unregulated growth, tissue invasion, and the ability to metastasize is characteristic of *malignant* neoplasms. Cancers are named based on their origin: those derived from epithelial tissue are called *carcinomas*, those derived from mesenchymal tissues are *sarcomas*, and those derived from hematopoietic tissue are *leukemias*, *lymphomas*, and *plasma cell dyscrasias* (including *multiple myeloma*).

Cancers nearly always arise as a consequence of genetic alterations, the vast majority of which begin in a single cell and therefore are monoclonal in origin. However, because a wide variety of genetic and epigenetic changes can occur in different cells within malignant tumors over time, most cancers are characterized by marked heterogeneity in the populations of cells. In addition, extrinsic factors in the cancer environment (e.g., the stroma, infiltrating cells, various cell-to-cell interactions, spatial orientation, secreted factors, and availability of oxygen and nutrients) vary in different areas within the tumor or different metastases, compounding this heterogeneity. This heterogeneity significantly complicates the treatment of most cancers because it is likely that there are subsets of cells that will be resistant to therapy for a variety of reasons and will therefore survive and proliferate even if the majority of cells are killed.

A few cancers appear to, at least initially, be primarily driven by an alteration in a dominant gene that produces uncontrolled cell proliferation. Examples include chronic myeloid leukemia (*abl*), about half of melanomas (*braf*), Burkitt's lymphoma (*c-myc*), and subsets of lung adenocarcinomas (*egfr*, *alk*, *ros1*, *met*, *ret*, *braf*, and *ntrk*). Genes that can promote cell growth when altered are often called *oncogenes*. They were first identified as critical elements of viruses that cause animal tumors; it was subsequently found that the viral genes had normal counterparts with important functions in the cell and had been captured and mutated by viruses as they passed from host to host.

However, most human cancers are characterized by a multiple-step process involving many genetic abnormalities, each of which contributes to the loss of control of cell proliferation and differentiation and the acquisition of capabilities, such as tissue invasion, the ability to metastasize, and angiogenesis (development of new blood vessels required for tumor growth). These properties are not found in the normal adult cell from which the tumor is derived. Indeed, normal cells have a large number of safeguards against DNA damage (including multiple DNA repair and extensive DNA damage response mechanisms), uncontrolled proliferation, and invasion. Many cancers go through recognizable steps of progressively more abnormal phenotypes: hyperplasia, to adenoma, to dysplasia, to carcinoma *in situ*, to invasive cancer with the ability to metastasize (Table 72-1). For most cancers, these changes occur over a prolonged period of time, usually many years.

In most organs, only primitive undifferentiated cells are capable of proliferating and cells lose the capacity to proliferate as they differentiate and acquire functional capabilities. The expansion of the primitive cells (stem cells) is linked to some functional need in the host through receptors that receive signals from the local environment or through hormonal and other influences delivered by the vascular supply. In the absence of such signals, the cells are at rest. The signals that keep the primitive cells at rest remain incompletely understood. These signals must be environmental, based on the observations that a regenerating liver stops growing when it has replaced the portion that has been surgically removed after partial hepatectomy and regenerating bone marrow stops growing when the peripheral blood counts return to

to signals that would limit their growth in tightly regulated normal tissue homeostasis. Most human cancers usually become clinically detectable when a primary mass is approximately 1 cm in diameter—such a mass consists of about 10^9 cells. Often, patients present with tumors that are approximately 10^{10} cells. Although it varies by type of cancer and where the primary tumor and metastases are located, a lethal tumor burden is usually about 10^{12} – 10^{13} cells. If all malignant cells were dividing without any cell death at the time of diagnosis, most patients would reach a lethal tumor burden in a very short time. However, human tumors grow by Gompertzian kinetics—this means that not every daughter cell produced by a cell division is actively dividing. In addition, the overall growth rate of a tumor depends on differences between growth rates of different cells within the tumor and rate of cell loss. The growth fraction of a tumor declines with time, largely due to factors in the microenvironment. The growth fraction of the first malignant cell is 100%, and by the time a patient presents for medical care, the growth fraction is estimated to be <10%, although the fraction varies between different types of cancers and even different cancers of the same type in different individuals. This fraction is often similar to the growth fraction of normal bone marrow and normal intestinal epithelium, the most highly proliferative normal tissues in the human body, a fact that may explain the dose-limiting toxicities to these tissues of agents that target dividing cells.

The implication of these data is that the tumor is slowing its own growth over time. How does it do this? The tumor cells have multiple genetic lesions that tend to promote proliferation, yet by the time the tumor is clinically detectable, its capacity for proliferation has declined. Better understanding of how a tumor slows its own growth would provide important clues for better cancer treatment. A number of factors, including those in the tumor microenvironment, are known to contribute to the decreased proliferation of tumor cells over time *in vivo*. Some cells are hypoxic and have inadequate supply of nutrients and energy. Some have sustained too much genetic damage to complete the cell cycle but have lost the capacity to undergo apoptosis and therefore survive but do not proliferate. However, an important subset is not actively dividing but retains the capacity to divide and can start dividing again under certain conditions such as when the tumor mass is reduced by treatments leading to improved conditions in the tumor microenvironment favorable for cell proliferation. Just as the bone marrow increases its rate of proliferation in response to bone marrow-damaging agents, the tumor also seems to sense when tumor cell numbers have been reduced and can respond by increasing growth rate. However, the critical difference is that the marrow stops growing when it has reached its production goals, whereas tumors do not.

The ultimate structure and organization of an organ are based on a number of factors including growth, migration, elimination, and death of various cells; communication between cells to establish the correct architecture; competition between cells; and the composition of the extracellular matrix that is produced. In addition to normal cells stopping proliferation in an organ when that is appropriate, normal tissues have various mechanisms for eliminating cells both in the process of development as well as ongoing homeostasis of an organ. These include mechanical processes based on a number of factors including cell size, shape, and topology between cells that can determine cell fate as well as an active process of cell extrusion, which plays a major role in the elimination of both cells that are no longer needed by the organ and cells that are damaged and potentially dangerous (such as those with mutations that might be precursors for malignancy). The process of cell extrusion may depend on cell cycle arrest in the S phase; aberrations in this process may contribute to the metastatic process. A variety of processes, including major alterations in cell cycle control, apoptosis and other mechanisms of cell death, and uncontrolled cell signaling, all contribute to defects in appropriate cell extrusion contributing to the development of cancer.

Additional tumor cell vulnerabilities are likely to be detected when we learn more about how normal cells respond to “stop” signals from their environment, and why and how tumor cells and tissues fail to heed such signals.

■ CELL CYCLE CHECKPOINTS

The cell division cycle consists of four phases—G₁ (growth and preparation for DNA synthesis), S (DNA synthesis), G₂ (preparation to divide), and M (mitosis, cell division). Cells can also exit the cell cycle and be quiescent (G₀). Progression of a cell through the cell cycle is tightly regulated at a number of checkpoints (especially at the G₁/S boundary, the G₂/M boundary, and during M [spindle checkpoint]) by an array of genes that are targeted by specific genetic alterations in cancer. Critical proteins in these control processes that are frequently mutated or otherwise inactivated in cancers are called tumor-suppressor genes. Examples include p53 and Rb (discussed below). In the first phase, G₁, preparations are made to replicate the genetic material. The cell stops before entering the DNA synthesis phase, or S phase, to take inventory. Are we ready to replicate our DNA? Is the DNA repair machinery in place to fix any mutations that are detected? Are the DNA replicating enzymes available? Is there an adequate supply of nucleotides? Is there sufficient energy to proceed? The retinoblastoma protein, Rb, plays a central role in placing a brake on the process until the cell is ready. When the cell determines that it is prepared to move ahead, sequential activation of cyclin-dependent kinases (CDKs) results in the inactivation of the brake, Rb, by phosphorylation. Phosphorylated Rb releases the S phase-regulating transcription factor, E2F/DP1, and genes required for S-phase progression are expressed. If the cell determines that it is unready to move ahead with DNA replication, a number of inhibitors are capable of blocking the action of the CDKs, including p21Cip2/Waf1, p16Ink4a, and p27Kip1. Nearly every cancer has one or more defects in the G₁ checkpoint that permit progression to S phase despite abnormalities in DNA repair machinery or other deficiencies that would affect normal DNA synthesis.

At the end of the G₂ phase and prior to the M phase, after the cell has exactly duplicated its DNA content, a second inventory is taken at the G₂ checkpoint. Have all of the chromosomes been fully duplicated? Were all segments of DNA copied only once? Has all damaged DNA been repaired? Do we have the right number of chromosomes and the right amount of DNA? If so, the cell proceeds to prepare for division by synthesizing mitotic spindle and other proteins needed to produce two daughter cells. If DNA damage is detected, the p53 pathway is normally activated. Called the guardian of the genome, p53 is a transcription factor that is normally present in the cell in very low levels. This level is generally regulated through its rapid turnover. Normally, p53 is bound to mdm2, a ubiquitin ligase that both inhibits p53 transcriptional activation and also targets p53 for degradation in the proteasome. When DNA damage is sensed, the ATM (ataxia-telangiectasia mutated) pathway is activated; ATM phosphorylates mdm2, releasing it from its inhibitory bond to p53. p53 then stops cell cycle progression, directs the synthesis of repair enzymes, or if the damage is too great, initiates apoptosis (programmed cell death) of the cell to prevent the propagation of a damaged cell (Fig. 72-1).

A second method of activating p53 involves the induction of p14ARF by hyperproliferative signals from oncogenes. p14ARF competes with p53 for binding to mdm2, allowing p53 to escape the effects of mdm2 and accumulate in the cell. p53 then stops cell cycle progression by activating CDK inhibitors such as p21 and/or initiating the apoptosis pathway. Not surprisingly given its critical role in controlling cell cycle progression, mutations in the gene for p53 on chromosome 17p are among the most frequent mutations in human cancers, although percentages vary between different cancers. Most commonly these mutations are acquired in the malignant tissue in one allele and the second allele is inactivated (such as by deletion or epigenetic silencing), leaving the cell unprotected from DNA-damaging agents or activated oncogenes. Some environmental exposures produce signature mutations in p53; for example, aflatoxin exposure leads to mutation of arginine to serine at codon 249 and leads to hepatocellular carcinoma. In rare instances, p53 mutations are in the germline (Li-Fraumeni syndrome) and produce a familial cancer syndrome. Another mechanism for inactivation of p53 in malignant cells is due to alterations in regulators such as overexpression of the inhibitory mdm2 protein. Whether inactivated by mutation or inhibited by regulatory factors, absence of normal p53 function leads to chromosomal instability and accumulation of DNA

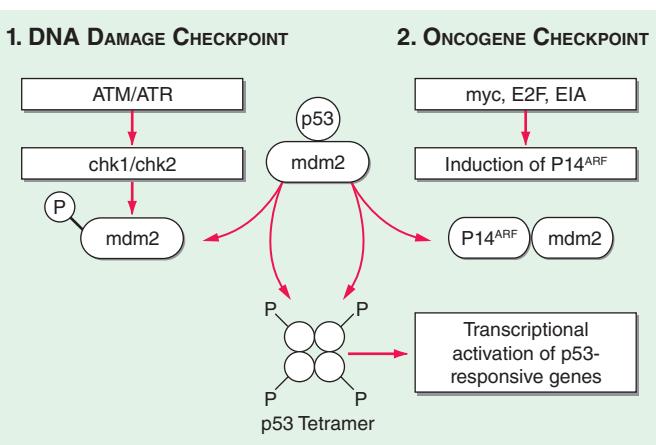


FIGURE 72-1 Induction of p53 by the DNA damage and oncogene checkpoints.

In response to noxious stimuli, p53 and mdm2 are phosphorylated by the ataxiatelangiectasia mutated (ATM) and related ATR serine/threonine kinases, as well as the immediate downstream checkpoint kinases, Chk1 and Chk2. This causes dissociation of p53 from mdm2, leading to increased p53 protein levels and transcription of genes leading to cell cycle arrest ($p21^{Cip1/Waf1}$) or apoptosis (e.g., the proapoptotic Bcl-2 family members Noxa and Puma). Inducers of p53 include hypoxemia, DNA damage (caused by ultraviolet radiation, gamma irradiation, or chemotherapy), ribonucleotide depletion, and telomere shortening. A second mechanism of p53 induction is activated by oncogenes such as *Myc*, which promote aberrant G₁/S transition. This pathway is regulated by a second product of the Ink4a locus, p14^{ARF} (p19 in mice), which is encoded by an alternative reading frame (ARF) of the same stretch of DNA that codes for p16^{Ink4a}. Levels of ARF are upregulated by *Myc* and E2F, and ARF binds to mdm2 and rescues p53 from its inhibitory effect. This *oncogene checkpoint* leads to the death or senescence (an irreversible arrest in G₁ of the cell cycle) of renegade cells that attempt to enter S phase without appropriate physiologic signals. Senescent cells have been identified in patients whose premalignant lesions harbor activated oncogenes, for instance, dysplastic nevi that encode an activated form of *BRAF* (see below), demonstrating that induction of senescence is a protective mechanism that operates in humans to prevent the outgrowth of neoplastic cells.

damage including acquisition of properties that give the abnormal cell a proliferative and survival advantage. Like Rb dysfunction, most cancers have mechanisms that disable the p53 pathway. Indeed, the importance of p53 and Rb in the development of cancer is underscored by the neoplastic transformation mechanism of human papillomavirus. This virus has two main oncogenes, E6 and E7. E6 acts to increase the rapid turnover of p53, and E7 acts to inhibit Rb function; inhibition of these two targets is required for transformation of epithelial cells.

Another cell cycle checkpoint exists when the cell is undergoing division (M phase); this is the spindle checkpoint, which acts to ensure that there is proper attachment of chromosomes to the mitotic spindle before progression through the cell cycle can occur. If the spindle apparatus does not properly align the chromosomes for division, if the chromosome number is abnormal (i.e., greater or less than $4n$), or if the centromeres are not properly paired with their duplicated partners, then the cell initiates a cell death pathway to prevent the production of aneuploid progeny (having an altered number of chromosomes). Abnormalities in the spindle checkpoint facilitate the development of aneuploidy, which is frequently found in cancers. In some tumors, aneuploidy is a predominant genetic feature.

In other tumors, a defect in the cells' ability to repair errors in the DNA, such as due to mutations in genes coding for the proteins critical for mismatched DNA repair, is the primary genetic lesion. Cancer cells can have defects in any of several DNA repair pathways in addition to mismatch repair, including deficient interstrand cross-link, double-strand breaks (homologous recombination or nonhomologous end joining repair), single-strand breaks, base excision, nucleotide excision, and translesional synthesis.

In general, tumors have either defects in chromosome number or defective DNA repair pathways but not both. Defects that lead to cancer include abnormal cell cycle checkpoints, inadequate DNA repair, and failure to preserve genome integrity leading to DNA damage. These

defects and the stress of the resultant increased DNA damage make cancer cells more vulnerable to additional DNA damage, which can be exploited by chemotherapy, radiation therapy, targeted therapy, and immunotherapy—the major systemic therapeutic approaches effective against cancer.

Alternatively, research is ongoing in an attempt to therapeutically restore the defects in cell cycle regulation and DNA repair that characterize cancer, although this remains a challenging problem because it is much more difficult to restore normal biologic function than to inhibit abnormal function of proteins driving cell proliferation, such as occurs with oncogenes. Newer approaches to gene editing (e.g., clustered regularly interspaced short palindromic repeats [CRISPR]) and subsequent modifications to this approach should make this more feasible.

CELLULAR SENESCENCE

The irreversible cessation of growth of normal cells while the cells remain viable has been termed cellular senescence. This was initially identified by the fact that when normal cells are placed in culture in vitro, most are not capable of sustained growth. They quickly reach a point where they either undergo cell death due to excessive DNA damage or other factors or they become senescent. Fibroblasts are an exception to this rule. When they are cultured, fibroblasts may divide 30–50 times and then they undergo what has been termed a “crisis” during which the majority of cells stop dividing (usually due to an increase in p21 expression, a CDK inhibitor). This form of senescence is termed replicative senescence. Many other cells die, and a small fraction emerge that have acquired genetic and epigenetic changes that permit their uncontrolled growth.

Among the cellular changes during in vitro propagation is telomere shortening. DNA polymerase is unable to replicate the tips of chromosomes, resulting in the loss of DNA at the specialized ends of chromosomes (called *telomeres*) with each replication cycle. At birth, human telomeres are 15- to 20-kb pairs long and are composed of tandem repeats of a six-nucleotide sequence (TTAGGG) that associate with specialized telomere-binding proteins to form a T-loop structure that protects the ends of chromosomes from being mistakenly recognized as damaged. The loss of telomeric repeats with each cell division cycle causes gradual telomere shortening, leading to growth arrest when one or more critically short telomeres trigger a p53-regulated DNA-damage checkpoint response. Cell death usually ensues when the unprotected ends of chromosomes lead to chromosome fusions or other catastrophic DNA rearrangements. Cells with certain abnormalities, such as those with nonfunctional pRb and p53, can bypass this growth arrest. *The ability to bypass telomere-based growth limitations is thought to be a critical step in the evolution of most malignancies.* This occurs by reactivation of telomerase expression in cancer cells. Telomerase is an enzyme that adds TTAGGG repeats onto the 3' ends of chromosomes. It contains a catalytic subunit with reverse transcriptase activity (hTERT) and an RNA component that provides the template for telomere extension. Most normal somatic cells do not express sufficient telomerase to prevent telomere attrition with each cell division. Exceptions include stem cells (such as those found in hematopoietic tissues, gut and skin epithelium, and germ cells) that require extensive cell division to maintain tissue homeostasis. More than 90% of human cancers express high levels of telomerase that prevent telomere shortening to critical levels and allow indefinite cell proliferation. In vitro experiments indicate that inhibition of telomerase activity leads to tumor cell apoptosis. Major efforts are underway to develop methods to inhibit telomerase activity in cancer cells. For example, the protein component of telomerase (hTERT) may act as one of the most widely expressed tumor-associated antigens and can be targeted by vaccine approaches. However, a caveat to targeting telomerase for anticancer treatment is the potential for inhibiting its activity in certain normal cells (such as stem cells) required for maintaining the normal physiologic state.

Although most of the functions of telomerase relate to cell division, it also has several other effects including interfering with the differentiated functions of at least certain stem cells. However, the impact on differentiated function of normal nonstem cells is less clear. The picture is further complicated by the fact that rare genetic defects in the

telomerase enzyme seem to cause dyskeratosis congenita (characterized by abnormalities in various rapidly dividing cells in the body including skin, nails, oral mucosa, hair, and bone marrow with increased risk for leukemia and certain other cancers). This can be associated with a number of other abnormalities including pulmonary fibrosis, bone marrow failure (aplastic anemia), or liver fibrosis. However, paradoxically, defects in nutrient absorption in the gastrointestinal tract, a site that should be highly sensitive to defective cell proliferation, are not seen. Much remains to be learned about how telomere shortening and telomere maintenance are related to human illness in general and cancer in particular.

A variety of other stresses on cells (both environmental and intrinsic including radiation, chemotherapy, reactive oxygen species, and oncogenic mutations) can also lead to senescence, primarily those that induce DNA damage similar to that seen in cells with shortened telomeres. This is termed *replicative senescence*.

SIGNAL TRANSDUCTION PATHWAYS IN CANCER CELLS

Signals that affect cell behavior come from adjacent cells, the stroma in which the cells are located, hormonal signals that originate remotely, and the cells themselves (autocrine signaling). These signals generally exert their influence on the receiving cell through activation of signal transduction pathways that have as their end result the induction of activated transcription factors that mediate a change in cell behavior or function or the acquisition of effector machinery to accomplish a new task. Although signal transduction pathways can lead to a wide variety of outcomes, many such pathways rely on cascades of signals that sequentially activate different proteins or glycoproteins and lipids or glycolipids, and the activation steps often involve the addition or removal of one or more phosphate groups on a downstream target. Other chemical changes can result from signal transduction pathways, but reversible phosphorylation and dephosphorylation play a major role. Proteins that add phosphate groups to other molecules (proteins, lipids, or nucleic acids) are called kinases. Two major classes of kinases involved in signal transduction pathways important for cancer cells are tyrosine kinases that phosphorylate tyrosine and serine/threonine kinases that phosphorylate serine/threonine either directly or indirectly. However, some kinases can phosphorylate both, such as the MEK kinases that can phosphorylate both threonine and tyrosine. Phosphatases (protein tyrosine phosphatases and protein serine/threonine phosphatases) remove the phosphate groups to reverse the kinase activity.

Various kinases play critical roles in signal transduction pathways important for malignant cells. These include a number of receptor tyrosine kinases (RTKs) as well as various protein kinases (both tyrosine and serine/threonine kinases) downstream of receptors that transmit the signals within the cell (Fig. 72-2). Two important signaling pathways are the RAS-RAF-MEK-ERK pathway and the phosphatidylinositol-3-kinase (PI3K) pathway (Fig. 72-2). Although pathways are depicted as distinct, there are complex interactions between pathways within cells.

Normally, kinase activity is short-lived and reversed by protein phosphatases. However, in many human cancers, RTKs or components of their downstream pathways are activated by mutation, gene amplification, or chromosomal translocations to have enhanced and/or prolonged activity. Because these pathways are important in regulating proliferation, survival, migration, and angiogenesis, they have been identified as important targets for cancer therapeutics.

Inhibition of kinase activity is effective in the treatment of a number of neoplasms. Lung cancers with mutations in the epidermal growth factor receptor are highly responsive to osimertinib as well as other inhibitors (Table 72-2). Inhibitors have been developed to treat lung cancers with other tyrosine kinase-activating mutations (including anaplastic lymphoma kinase [ALK], ROS1, NTRK, MET, and RET). BRAF (a serine/threonine kinase) inhibitors are highly effective in melanomas and thyroid cancers and are also used in combination with other agents for lung and colon cancers in which BRAF is mutated. Targeting the MEK protein (which phosphorylates both threonine and

tyrosine residues) downstream of BRAF also has activity against BRAF mutant melanomas, and combined inhibition of BRAF and MEK is more effective than either alone with activity that extends to BRAF mutant lung cancer. Janus kinase (JAK) inhibitors are active in myeloproliferative syndromes in which JAK2 activation is a pathogenetic event. Imatinib (which targets a number of tyrosine kinases) is an effective agent in tumors that have translocations of the c-Abl and BCR gene (such as chronic myeloid leukemia), mutant c-Kit (gastrointestinal stromal cell tumors), or mutant platelet-derived growth factor receptor (PDGFR α ; gastrointestinal stromal tumors). Second-generation inhibitors of BCR-Abl, dasatinib and nilotinib, are even more effective, and the third-generation agent bosutinib has activity in some patients who have progressed on other inhibitors, while the third-generation agent ponatinib has activity against the T315I mutation, which is resistant to the other agents. Although almost all tyrosine kinase inhibitors are not entirely selective for one protein, certain inhibitors have significant activity against a broad number of proteins. These include sorafenib, regorafenib, cabozantinib, sunitinib, and lenvatinib. These have shown antitumor activity in various malignancies, including renal cell cancer (RCC) (sorafenib, sunitinib, cabozantinib, lenvatinib), hepatocellular carcinoma (sorafenib, regorafenib, lenvatinib), gastrointestinal stromal tumor (GIST) (sunitinib, regorafenib), thyroid cancer (sorafenib, cabozantinib, lenvatinib), colorectal cancer (regorafenib), and pancreatic neuroendocrine tumors (sunitinib).

Inhibitors of the PI3K pathway also have been approved for cancer therapy. The PI3K family includes three classes and several isoforms within each class. Inhibitors against different isoforms have proved effective against different types of malignancies, with inhibitors of the delta isoform (either specifically or also with inhibition of other isoforms; e.g., idelalisib) having activity against lymphoid malignancies, whereas the specific inhibitor of a mutation in the alpha isoform (alpelisib) has activity against breast cancers with this mutation. Inhibitors of mammalian target of rapamycin (mTOR; which is downstream of PI3K; e.g., everolimus, temsirolimus) are active in RCC, pancreatic neuroendocrine tumors, and breast cancer. Additional inhibitors of the PI3K pathway and other phospholipid signaling pathways such as the phospholipase C-gamma pathway, which are involved in a large number of cellular processes important in cancer development and progression, are being evaluated.

The list of active agents and treatment indications is growing rapidly (Table 72-2). These agents have ushered in a new era of personalized therapy. It is becoming more common for tumor biopsies to be assessed for specific molecular changes that predict response and to have clinical decision-making guided by those results. This is now an important component of standard therapy for metastatic lung, gastroesophageal, melanoma, breast, and colorectal cancers as well as in adjuvant therapy for breast cancer.

An alternative approach to testing samples directly from tumors is to test blood for the presence of mutations or amplification in circulating tumor DNA, which has the significant advantage of being noninvasive. As cancers grow, some of the cells die and break down with release of cellular contents, including DNA, into the circulation. Sensitive methods have been developed to detect this DNA and to identify mutations and other DNA changes in the malignant cells. This has the potential advantage over tumor biopsies of sampling all of the tumor and not being limited to one site that may not be representative of the overall tumor heterogeneity. In addition to identifying potential changes that can be targeted for therapy, there is also the potential for monitoring a patient's response to therapy, identifying resistance mechanisms to therapy earlier, detecting disease recurrence before it can be detected by tumor markers or scans, monitoring bodily fluids in addition to blood, and possibly providing a means of earlier initial detection of cancer if sufficiently sensitive and specific detection methods can be developed.

However, none of these targeted therapies has yet been curative by themselves for any malignancy, although prolonged periods of disease control lasting many years frequently occur in chronic myeloid leukemia (CML), including a >80% survival rate at 10 years. The reasons for the failure to cure are not completely defined, although resistance

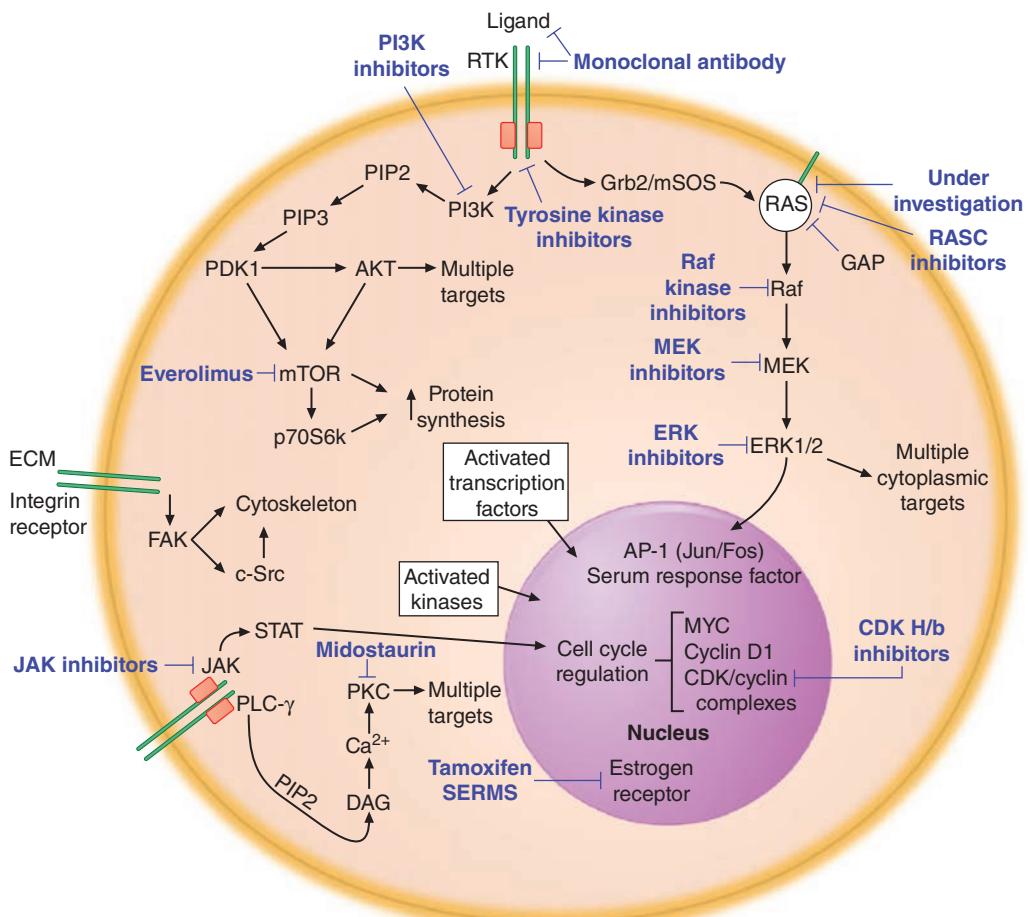


FIGURE 72-2 Therapeutic targeting of signal transduction pathways in cancer cells. Three major signal transduction pathways are activated by receptor tyrosine kinases (RTKs). 1. The protooncogene Ras is activated by the Grb2/mSOS guanine nucleotide exchange factor, which induces an association with Raf and activation of downstream kinases (MEK and ERK1/2). 2. Activated PI3K phosphorylates the membrane lipid PIP₂ to generate PIP₃, which acts as a membrane-docking site for a number of cellular proteins including the serine/threonine kinases PDK1 and Akt. PDK1 has numerous cellular targets, including Akt and mTOR. Akt phosphorylates target proteins that promote resistance to apoptosis and enhance cell cycle progression, while mTOR and its target p70S6K upregulate protein synthesis to potentiate cell growth. 3. Activation of PLC- γ leads to the formation of diacylglycerol (DAG) and increased intracellular calcium, with activation of multiple isoforms of PKC and other enzymes regulated by the calcium/calmodulin system. Other important signaling pathways involve non-RTKs that are activated by cytokine or integrin receptors. Janus kinases (JAK) phosphorylate STAT (signal transducer and activator of transcription) transcription factors, which translocate to the nucleus and activate target genes. Integrin receptors mediate cellular interactions with the extracellular matrix (ECM), inducing activation of FAK (focal adhesion kinase) and c-Src, which activate multiple downstream pathways, including modulation of the cell cytoskeleton. Many activated kinases and transcription factors migrate into the nucleus, where they regulate gene transcription, thus completing the path from extracellular signals, such as growth factors, to a change in cell phenotype, such as induction of differentiation or cell proliferation. The nuclear targets of these processes include transcription factors (e.g., Myc, AP-1, and serum response factor) and the cell cycle machinery (cyclin-dependent kinases [CDKs] and cyclins). Inhibitors of many of these pathways have been developed for the treatment of human cancers. Examples of inhibitors that are either approved or are currently being evaluated in clinical trials are shown in purple type.

to the treatment ultimately develops in most patients. In some tumors, resistance to kinase inhibitors is related to proliferation of cells with a mutation in the target kinase that inhibits drug binding. Many of these kinase inhibitors act as competitive inhibitors of the ATP-binding pocket. ATP is the phosphate donor in these phosphorylation reactions. For example, mutation in the critical BCR-ABL kinase in the ATP-binding pocket (such as the threonine to isoleucine change at codon 315 [T315I]) can prevent imatinib binding. Other resistance mechanisms include alterations in other signal transduction pathways to bypass the inhibited pathway. As resistance mechanisms continue to be better defined, rational strategies to overcome resistance are emerging. In addition, many kinase inhibitors are less specific for an oncogenic target than was hoped, and toxicities related to off-target inhibition of kinases limit the use of the agent at a dose that would optimally inhibit the cancer-relevant kinase.

Antibodies against protein targets more highly expressed on malignant than normal cells can also be used to deliver highly toxic compounds relatively specifically to cancer cells. Examples of protein targets for currently approved antibody-drug conjugates include CD30 for Hodgkin's and anaplastic lymphomas; HER2 on breast cancer; CD33 on

acute myeloid leukemias; CD22 on B-cell acute lymphocytic and hairy cell leukemias; and CD79b on diffuse large B-cell lymphomas.

Another strategy to enhance the antitumor effects of targeted agents is to use them in rational combinations with each other as well as with chemotherapy or immunotherapy agents that kill cells in ways distinct from agents targeting specific mutant or overexpressed proteins. Combinations of trastuzumab (a monoclonal antibody that targets the HER2 receptor [member of the EGFR family]) with chemotherapy have significant activity against breast and stomach cancers that have high levels of expression of the HER2 protein. The activity of trastuzumab and chemotherapy can be enhanced further by combinations with another targeted monoclonal antibody (pertuzumab), which prevents dimerization of the HER2 receptor with other HER family members including HER3.

Although targeted therapies have not yet resulted in cures when used alone, their use in the adjuvant setting and when combined with other effective treatments has substantially increased the fraction of patients cured. For example, the addition of rituximab, an anti-CD20 antibody, to combination chemotherapy in patients with diffuse large B-cell lymphoma improves cure rates by ~15%. The addition

TABLE 72-2 Some FDA-Approved Molecularly Targeted Agents for the Treatment of Cancer

DRUG	MOLECULAR TARGET	DISEASE	MECHANISM OF ACTION
All- <i>trans</i> retinoic acid	PML-RAR α oncogene	Acute promyelocytic leukemia M3 AML, t(15;17)	Inhibits transcriptional repression by PML-RAR α
Imatinib	Bcr-Abl, c-Abl, c-Kit, PDGFR- α/β	Chronic myeloid leukemia, GIST	Blocks ATP binding to tyrosine kinase active site
Ripretinib	c-Kit, PDGFR- α	GIST	Inhibits tyrosine kinase activity
Dasatinib, nilotinib, ponatinib, bosutinib	Bcr-Abl (primarily)	Chronic myeloid leukemia	Blocks ATP binding to tyrosine kinase active site
Sunitinib	c-Kit, VEGFR-2, PDGFR- β , Flt-3	GIST, RCC, PNET	Inhibits activated c-Kit and PDGFR in GIST; inhibits VEGFR in RCC and probably in PNET
Sorafenib	RAF, VEGFR-2, PDGFR- α/β , Flt-3, c-Kit	RCC, hepatocellular carcinoma (HCC), differentiated thyroid cancer, desmoid	Targets VEGFR pathways in RCC and HCC. Possible activity against BRAF in thyroid cancer
Regorafenib	VEGFR1–3, TIE-2, FGFR1, KIT, RET, PDGFR	Colorectal cancer, GIST, HCC	Competitive inhibitor ATP binding site of tyrosine kinase domain multiple kinases including VEGFR
Larotrectinib, entrectinib	NTRK	Cancers with NTRK mutation	Competitive inhibitor of ATP binding site of the tyrosine kinase domain of NTRK
Axitinib	VEGFR1–3	RCC	Competitive inhibitor ATP binding site of tyrosine kinase domain VEGF receptors
Erlotinib	EGFR	NSCLC, pancreatic cancer	Competitive inhibitor of the ATP-binding site of the EGFR
Afatinib	EGFR (and other HER family)	NSCLC	Irreversible inhibitor of ATP-binding site of HER family members
Osimertinib	EGFR (T790M)	NSCLC	Inhibits EGFR mutations including T790M mutant NSCLC
Dacomitinib	EGFR	NSCLC (exon19 deletion/exon 21 L858R)	Inhibits EGFR mutant lung cancer
Erdafitinib, pemigatinib	FGFR2, FGFR3	Urothelial (erdafitinib), cholangiocarcinoma (pemigatinib)	Inhibits tyrosine kinase of FGFR
Lapatinib, tucatinib	HER2/neu	Breast cancer	Competitive inhibitor of the ATP-binding site of HER2
Crizotinib, ceritinib, alectinib, brigatinib, lorlatinib	ALK	NSCLC	Inhibitor of ALK tyrosine kinase
Crizotinib, entrectinib	ROS1	NSCLC	Inhibitor of ROS1 tyrosine kinase
Palbociclib, ribociclib, abemaciclib	CDK4/6	Breast	Inhibitor of CDK4/6
Bortezomib, carfilzomib, ixazomib	Proteasome	Multiple myeloma	Inhibits proteolytic degradation of multiple cellular proteins
Vemurafenib, dabrafenib	BRAF	Melanoma	Inhibitor of serine-threonine kinase domain of V600E mutant of BRAF
Encorafenib	MEK	CRC	Inhibits BRAFV600E mutation; used in combination with cetuximab
Trametinib, Cobimetinib	MEK	Melanoma	Inhibitor of serine-threonine kinase domain of MEK
Cabozantinib	RET, MET, VEGFR	MTC, RCC	Competitive inhibitor of ATP-binding site of tyrosine kinase domain of multiple kinases, including VEGFR2 and RET
Capmatinib	MET	NSCLC with MET exon14 deletions	
Vandetanib	RET, VEGFR, EGFR	MTC	Competitive inhibitor of ATP-binding site of tyrosine kinase domain of multiple kinases, including RET
Selpercatinib	RET	NSCLC, MTC, RET fusion thyroid cancer	Inhibitor of RET, VEGFR1, VEGFR2 tyrosine kinases
Temsirolimus	mTOR	RCC	Competitive inhibitor of mTOR serine-threonine kinase
Everolimus	mTOR	RCC, PNET	Binds to immunophilin FK binding protein-12, which forms a complex that inhibits mTOR kinase
Vorinostat, romidepsin, belinostat	HDAC	CTCL/PTL	HDAC inhibitor, epigenetic modulation
Panobinostat	HDAC	MM	HDAC inhibitor, epigenetic modulation
Ruxolitinib	JAK-1, 2	Myelofibrosis	Competitive inhibitor of tyrosine kinase
Vismodegib	Hedgehog pathway	Basal cell cancer (skin)	Inhibits smoothened in hedgehog pathway
Lenvatinib	Multikinase inhibitor (VEGFR, FGFR, PGFR- α , others)	RCC, thyroid cancer, HCC	Competitive inhibitor of ATP-binding site of tyrosine kinase domain of multiple kinases
Olaparib, rucaparib, niraparib, talazoparib	PARP	BRCA mutant ovarian, breast, prostate, pancreas cancers; not all agents approved for all cancers	Inhibits PARP and DNA repair
Venetoclax	BCL-2	CLL (with 17p deletion)	Inhibits BCL-2 and enhances apoptosis
Ibrutinib, acalabrutinib	Bruton tyrosine kinase (BTK)	CLL, MCL, MZL, SLL, WM	Inhibitor of BTK

(Continued)

TABLE 72-2 Some FDA-Approved Molecularly Targeted Agents for the Treatment of Cancer (Continued)

DRUG	MOLECULAR TARGET	DISEASE	MECHANISM OF ACTION
Ivosidenib	IDH1	AML	IDH1 inhibitor
Gilteritinib	FLT3	AML	FLT3 inhibitor
Idelalisib	PI3K-delta	CLL, SLL, FL	Inhibits PI3K-delta, preventing proliferation and inducing apoptosis
Alpelisib	PIK3CA	Breast cancer with a PIK3CA mutation	Inhibits PIK3CA
Monoclonal Antibodies Alone			
Trastuzumab	HER2/neu (ERBB2)	Breast cancer, gastric cancer	Binds HER2 on tumor cell surface and induces receptor internalization
Pertuzumab	HER2/neu (ERBB2)	Breast cancer	Binds HER2 on tumor cell surface at distinct site from trastuzumab and prevents binding to other receptors
Cetuximab	EGFR	Colon cancer, squamous cell carcinoma of the head and neck	Binds extracellular domain of EGFR and blocks binding of EGF and TGF- α ; induces receptor internalization. Potentiates the efficacy of chemotherapy and radiotherapy
Panitumumab	EGFR	Colon cancer	Similar to cetuximab but fully humanized rather than chimeric
Necitumumab	EGFR	Squamous NSCLC	Binds EGFR
Rituximab	CD20	B-cell lymphomas and leukemias that express CD20	Multiple potential mechanisms, including direct induction of tumor cell apoptosis and immune mechanisms
Alemtuzumab	CD52	Chronic lymphocytic leukemia and CD52-expressing lymphoid tumors	Immune mechanisms
Bevacizumab	VEGF	Colorectal, lung cancers, RCC, glioblastoma	Inhibits angiogenesis by high-affinity binding to VEGF
Ziv-aflibercept	VEGFA, VEGFB, PLGF	Colorectal cancers	Inhibits angiogenesis by high-affinity binding to VEGFA, VEGFB, and PLGF
Ramucirumab	VEGFR	Gastric, colorectal, lung cancers	Inhibits angiogenesis by binding to VEGFR
Ipilimumab	CTLA-4	Melanoma, HCC, MSI-high colorectal cancer	Blocks CTLA-4, preventing interaction with CD80/86 and T-cell inhibition
Nivolumab, pembrolizumab	PD-1	Melanoma, head and neck cancer, NSCLC, SCLC, Hodgkin's disease, urothelial cancer, RCC, HCC, gastric cancer, MSI-high cancers, endometrial cancer	Blocks PD-1, preventing interaction with PD-L1 and T-cell inhibition
Atezolizumab, durvalumab, avelumab	PD-L1	NSCLC, urothelial cancer, SCLC (durvalumab), HCC (atezolizumab), Merkel cell cancer (avelumab)	Blocks PD-L1, preventing interaction with PD-1 and T-cell inhibition
Denosumab	Rank ligand	Breast, prostate	Inhibits Rank ligand, primary signal for bone removal
Dinutuximab	Glycolipid GD2	Neuroblastoma (pediatric)	Immune-mediated attack on GD2-expressing cells
Daratumumab	CD38	MM	Binds to CD38 on MM cells causing apoptosis by antibody-dependent or complement-mediated cytotoxicity
Elotuzumab	SLAMF7	MM	Activating NK cells to kill MM cells
Olaratumab	PDGFR α	Soft tissue sarcomas	Blocks PDGFR α activity
Blinatumomab	CD19 and CD3	Ph-relapsed precursor B-cell ALL	Binds CD19 on ALL cells and CD3 on T cells; immune attack on CD19-expressing cells
Antibody-Chemotherapy Conjugates			
Brentuximab vedotin	CD30	Hodgkin's disease, anaplastic lymphoma	Delivery of chemotherapeutic agent (MMAE) to CD30-expressing tumor cells
Ado-trastuzumab emtansine	HER2	Breast cancer	Delivery of chemotherapeutic agent emtansine to HER2-expressing breast cancer cells
Fam-trastuzumab	HER2	Breast cancer, gastric cancer	Delivery of chemotherapeutic agent deruxtecan to HER2-expressing breast cancer cells
CAR-T Cells			
Tisagenlecleucel, axicabtagene ciloleucel	CD19	ALL (tisagenlecleucel), DLBCL/high-grade BCL (acicabtagene ciloleucel)	Targeted T cells to protein on surface of malignant cells

Abbreviations: ALL, acute lymphocytic leukemia; AML, acute myeloid leukemia; BCL, B-cell lymphoma; CAR-T, chimeric antigen receptor T cells; CLL, chronic lymphocytic leukemia; CRC, colorectal cancer; CTCL, cutaneous T-cell lymphoma; DLBCL, diffuse large B-cell lymphoma; EGFR, epidermal growth factor receptor; FDA, U.S. Food and Drug Administration; FGFR, fibroblast growth factor receptor; FL, follicular lymphoma; Flt-3, fms-like tyrosine kinase-3; GIST, gastrointestinal stromal tumor; HDAC, histone deacetylases; MCL, mantle cell lymphoma; MM, multiple myeloma; MSI, microsatellite instability; MTC, medullary thyroid cancer; mTOR, mammalian target of Rapamycin; MZL, mantle zone lymphoma; NK, natural killer; NSCLC, non-small-cell lung cancer; PARP, poly-ADP ribose polymerase; PDGFR, platelet-derived growth factor receptor; PLGF, placenta growth factor; PML-RAR α , promyelocytic leukemia-retinoic acid receptor-alpha; PNET, pancreatic neuroendocrine tumors; PTL, peripheral T-cell lymphoma; RCC, renal cell cancer; t(15;17), translocation between chromosomes 15 and 17; SCLC, small-cell lung cancer; SLL, small lymphocytic lymphoma; TGF- α , transforming growth factor-alpha; VEGFR, vascular endothelial growth factor receptor; WM, Waldenström's macroglobulinemia.

of trastuzumab, antibody to HER2, to combination chemotherapy in the adjuvant treatment of HER2-positive breast cancer significantly improves overall survival.

A major effort continues to develop targeted therapies for mutations in the *ras* family of genes, which play a critical role in transmitting signals through a number of downstream signaling pathways including the MAP (mitogen-activated protein) kinase and PI3K pathways. Mutations in *ras* are the most common mutations in oncogenes in cancers (especially *kras*) but have proved to be very difficult targets for a number of reasons related to the structure of RAS proteins as well as mechanisms of activation and inactivation (active when bound to guanosine triphosphate [GTP] and inactive when bound to guanosine diphosphate [GDP]). RAS proteins are not kinases but bind directly to the BRAF serine/threonine kinase with preferential binding when RAS is in the active GTP bound state. Preliminary evidence indicates antitumor activity of agents that target one of the mutant forms of KRAS (12C) that is found in a subset of cancers. Indirect inhibition of RAS function by inhibiting farnesyl transferase, which is important for RAS binding to the membrane and is required for activation, has shown some promise against HRAS mutant head and neck cancers. Targeted therapies against a subset of proteins downstream of RAS in the MAP kinase signaling pathway (including BRAF and MAP kinase) have proven to have significant antitumor activity against V600E *BRAF* mutant melanoma, with improved efficacy when they are used in combination. However, similar activity is not seen against *ras* mutant tumors. Additional targeted therapies against other proteins downstream of RAS (including ERK, or combinations of MAP kinase inhibitors and immunotherapy) are being studied, both individually and in combination. However, at this time, there is no clinically approved approach to inhibiting RAS mutant tumors.

One of the strategies for new drug development is to take advantage of so-called oncogene addiction. This situation (Fig. 72-3) is created when a tumor cell develops an activating mutation in an oncogene that becomes a dominant pathway for survival and growth with reduced contributions from other pathways, even when there may be abnormalities in those pathways. This dependency on a single pathway creates a cell that is vulnerable to inhibitors of that oncogene pathway. For example, cells harboring mutations in *BRAF* are very sensitive to MEK inhibitors that inhibit downstream signaling in the *BRAF* pathway.

Proteins critical for transcription of other proteins essential for malignant cell survival or proliferation provide another potential target for treating cancers. The transcription factor nuclear factor (NF)- κ B is a heterodimer composed of p65 and p50 subunits that associate with an inhibitor, I κ B, in the cell cytoplasm. In response to growth factor or cytokine signaling, a multi-subunit kinase called IKK (I κ B-kinase) phosphorylates I κ B and directs its degradation by the ubiquitin/proteasome system. NF- κ B, free of its inhibitor, translocates to the nucleus and activates target genes, many of which promote the survival of tumor cells. One of the mechanisms by which novel drugs called *proteasome inhibitors* are thought to produce an anticancer effect is by blocking the proteolysis of I κ B, thereby preventing NF- κ B activation.

For reasons that have not been fully elucidated, this has a differential toxicity effect on tumor, as compared to normal, cells. Although this mechanism appears to be an important aspect of the antitumor effects of proteasome inhibitors, there are other effects involving the inhibition of the degradation of multiple cellular proteins important in malignant cell survival or proliferation. Proteasome inhibitors (bortezomib, carfilzomib, ixazomib) have activity in patients with multiple myeloma, including partial and complete remissions. Inhibitors of IKK are also in development, with the hope of more selectively blocking the degradation of I κ B, thus “locking” NF- κ B in an inhibitory complex and rendering the cancer cell more susceptible to apoptosis-inducing agents. Many other transcription factors are activated by phosphorylation, which can be prevented by tyrosine or serine/threonine kinase inhibitors, a number of which are currently in clinical trials.

Estrogen receptors (ERs) and androgen receptors (ARs), members of the steroid hormone family of nuclear receptors, are targets of inhibition by drugs used to treat breast and prostate cancers, respectively. Selective estrogen receptor modulators (SERMs) have been developed as a treatment approach for ER-positive breast cancer. Tamoxifen, a partial agonist and antagonist of ER function, is frequently used in breast cancer and can mediate tumor regression in metastatic breast cancer and can prevent disease recurrence in the adjuvant setting. Tamoxifen binds to the ER and modulates its transcriptional activity, inhibiting activity in the breast but promoting activity in bone but unfortunately also in uterine epithelium, leading to a small increased risk of uterine cancer. Attempts have been made to develop SERMs that would have antiestrogenic effects in both breast and uterus while maintaining protective effects on bone. However, none of these to date has been an improvement over tamoxifen. Aromatase inhibitors, which

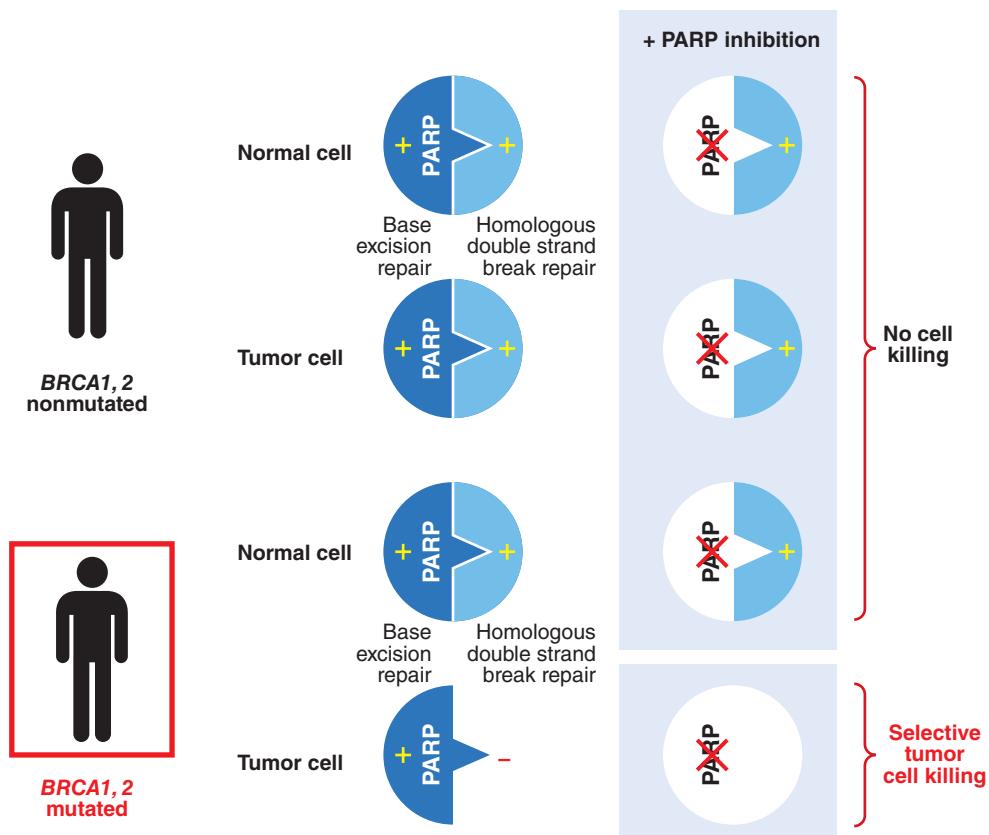


FIGURE 72-3 Synthetic lethality. Genes are said to have a synthetic lethal relationship when mutation of either gene alone is tolerated by the cell, but mutation of both genes leads to lethality, as originally noted by Bridges and later named by Dobzhansky. Thus, mutant *gene a* and *gene b* have a synthetic lethal relationship, implying that the loss of one gene makes the cell dependent on the function of the other gene. In cancer cells, loss of function of a DNA repair gene like *BRCA1*, which repairs double-strand breaks, makes the cell dependent on base excision repair mediated in part by *PARP*. If the *PARP* gene product is inhibited, the cell attempts to repair the break using the error-prone nonhomologous end-joining method, which results in tumor cell death. High-throughput screens can now be performed using isogenic cell line pairs in which one cell line has a defined defect in a DNA repair pathway. Compounds can be identified that selectively kill the mutant cell line; targets of these compounds have a synthetic lethal relationship to the repair pathway and are potentially important targets for future therapeutics.

block the conversion of androgens to estrogens in breast and subcutaneous fat tissues, have demonstrated improved clinical efficacy compared with tamoxifen in postmenopausal women and are often used as first-line therapy in postmenopausal patients with ER-positive disease. They are occasionally used in premenopausal patients with ER-positive disease in combination with ovarian suppression approaches such as luteinizing hormone-releasing hormone (LHRH) agonists. A number of approaches have been developed for blocking androgen stimulation of prostate cancer, including decreasing production by the testicles (e.g., orchectomy, LHRH agonists or antagonists), directly blocking actions of androgen (a number of agents have been developed to do this), or blocking production by inhibiting the enzyme CYP17, which is central in production of androgens from cholesterol (**Chap. 79**).

CANCER-SPECIFIC GENETIC CHANGES AND SYNTHETIC LETHALITY

The concepts of oncogene addiction and synthetic lethality have spurred new drug development targeting oncogene- and tumor-suppressor pathways. As discussed earlier in this chapter and outlined in Fig. 72-3, cancer cells can become dependent upon signaling pathways containing activated oncogenes; this can effect proliferation (i.e., mutated KRAS, BRAF, overexpressed MYC, or activated tyrosine kinases). Additional genetic changes in malignant cells or unique features of tumors including defects in DNA repair (e.g., loss of *BRCA1* or *BRCA2* gene function), modifications in cell cycle control (e.g., changes in protein levels or mutations in cyclins and cyclin-dependent kinases), enhanced survival mechanisms (overexpression of Bcl-2 or NF- κ B), altered cell metabolism (such as occurs when mutant KRAS enhances glucose uptake and aerobic glycolysis), tumor-stromal interactions, and angiogenesis (e.g., production of vascular endothelial growth factor [VEGF] in response to HIF-2 α in RCC) can also be successfully exploited to relatively specifically target cancers. However, resistance to inhibition of specific oncogenic pathways almost always eventually develops. In addition, targeting defects in tumor-suppressor genes has been much more difficult, both because the target of mutation is often deleted and because it is much more difficult to restore normal function than to inhibit abnormal function of a protein.

Synthetic lethality occurs when loss of function in either of two or more genes individually has limited effects on cell survival but loss of function in both (or more) genes leads to cell death. In the case of oncogene addicted pathways, identifying genes that have a synthetic lethal relationship with the activated pathway may allow enhanced cell killing and decreased resistance by targeting those genes or their proteins. In the case of mutant tumor-suppressor genes, identifying genes that have a synthetic lethal relationship to those mutated pathways may allow targeting by inhibiting proteins required uniquely by those cells for survival or proliferation (Fig. 72-3). This is a much more tractable approach than attempting to repair normal function of the mutant suppressor gene itself. Examples of synthetic lethality with clinical impact have been identified. For instance, cells with mutations in the *BRCA1* or *BRCA2* tumor-suppressor genes (e.g., a subset of breast and ovarian cancers) are unable to repair DNA damage by homologous recombination. Poly-ADP ribose polymerase (PARP) is a family of proteins important for single-strand break (SSB) DNA repair. PARP inhibition results in selective killing of cancer cells that have lost *BRCA1* or *BRCA2* function. A number of PARP inhibitors have been approved for treatment of ovarian, breast, prostate, and pancreatic cancers that have mutations in BRCA genes, as well as for maintenance therapy of ovarian cancer and are likely to have activity in other tumors with defective DNA repair mechanisms. The concept of synthetic lethality provides a framework for genetic screens to identify other synthetic lethal combinations involving known tumor-suppressor genes and development of novel therapeutic agents to target dependent pathways. Other unique aspects of malignant tumors, including those outlined elsewhere in the chapter, may also be vulnerable to synthetic lethal interactions.

■ EPIGENETIC INFLUENCES ON CANCER GENE TRANSCRIPTION

Chromatin structure regulates the hierarchical order of sequential gene transcription that governs differentiation and tissue homeostasis.

Disruption of chromatin remodeling (the process of modifying chromatin structure to control exposure of specific genes to transcriptional proteins, thereby controlling the expression of those genes) leads to aberrant gene expression that can significantly alter the biology of cells including inducing proliferation or migration of cells. *Epi*genetic changes are those that alter the pattern of gene expression that persist across at least one cell division, but are not caused by changes in the DNA code. These include alterations of chromatin structure mediated by methylation of cytosine residues of DNA (primarily in context of CpG dinucleotides in somatic cells), modification of histones by altering acetylation or methylation, or changes in higher-order chromosome structure (Fig. 72-4). Appropriate control of DNA methylation is essential for normal cell function and development, and both altered methylation and hypomethylation of histones occur in cancers. Hypermethylation of DNA promoter regions is a common mechanism by which tumor-suppressor loci are epigenetically silenced in cancer cells. Thus, one allele of a tumor-suppressor gene may be inactivated by mutation or deletion, while expression of the other allele is epigenetically silenced, usually by methylation, leading to loss of gene function. Aberrant hypomethylation is also frequently found in a number of cancers consistent with the dysregulated pattern of gene transcription that is a hallmark of cancer cells, with some genes being inappropriately turned off while others are inappropriately turned on.

Acetylation of the amino terminus of the core histones H3 and H4 induces an open chromatin conformation that promotes transcription initiation. Histone acetylases are components of coactivator complexes recruited to promoter/enhancer regions by sequence-specific transcription factors during the activation of genes (Fig. 72-4). Histone deacetylases (HDACs; multiple HDACs are encoded in the human genome) are recruited to genes by transcriptional repressors and prevent the initiation of gene transcription. Methylated cytosine residues in promoter regions become associated with methyl cytosine-binding proteins that recruit protein complexes with HDAC activity. The balance between permissive and inhibitory chromatin structure is therefore largely determined by the activity of transcription factors in modulating the “histone code” and the methylation status of the genetic regulatory elements of genes.

The pattern of gene transcription is aberrant in all human cancers, and in many cases, epigenetic events are responsible. Epigenetic events play a critical role in carcinogenesis (e.g., long-lasting changes in methylation induced by smoking) and are found in premalignant lesions. Unlike genetic events that alter DNA primary structure (e.g., deletions), epigenetic changes are potentially reversible and appear amenable to therapeutic intervention. In certain human cancers, including a subset of pancreatic cancers and multiple myeloma, the p16^{Ink4a} promoter is inactivated by methylation, thus permitting the unchecked activity of CDK4/cyclin D and rendering pRB nonfunctional. In sporadic forms of renal, breast, and colon cancer, the von Hippel-Lindau (VHL), breast cancer 1 (*BRCA1*), and serine/threonine kinase 11 (*STK11*) genes, respectively, can be epigenetically silenced. Other targeted genes include the p15^{Ink4b} CDK inhibitor, glutathione-S-transferase (which detoxifies reactive oxygen species [ROS]), and the E-cadherin molecule (important for junction formation between epithelial cells). Epigenetic silencing can affect genes involved in DNA repair, thus predisposing to further genetic damage. Examples include MLH1 (mutL homologue in sporadic colon cancers that have microsatellite instability) and MSH2 in a subset of hereditary nonpolyposis colon cancer patients who have a mutation in the 3' end of epithelial cell adhesion molecule (EPCAM). These are critical genes involved in repair of mismatched bases that occur during DNA synthesis, and their silencing can lead to mutations in the DNA.

Human leukemias often have chromosomal translocations that code for novel fusion proteins with activities that alter chromatin structure by interacting with HDACs or histone acetyl transferases (HATs). For example, the promyelocytic leukemia-retinoic acid receptor α (PML-RAR α) fusion protein, generated by the t(15;17) translocation observed in most cases of acute promyelocytic leukemia (APL), binds to promoters containing retinoic acid response elements and recruits HDACs to these promoters, effectively inhibiting gene expression.

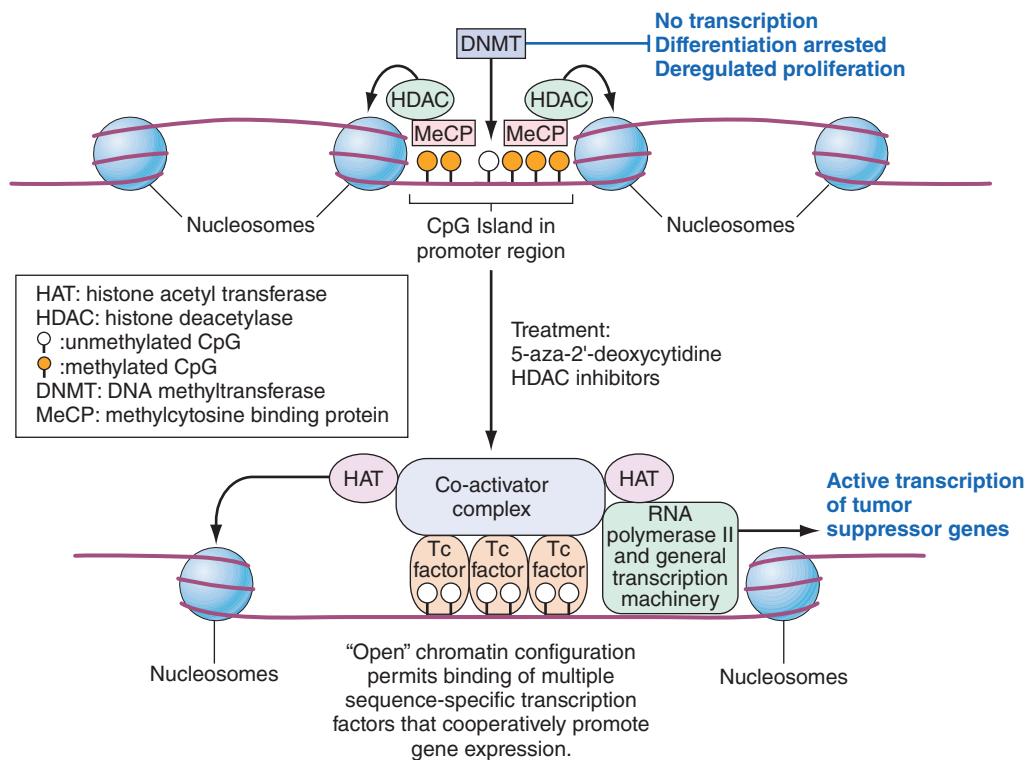


FIGURE 72-4 Epigenetic regulation of gene expression in cancer cells. Tumor-suppressor genes are often epigenetically silenced in cancer cells. In the upper portion, a CpG island within the promoter and enhancer regions of the gene has been methylated, resulting in the recruitment of methyl-cytosine binding proteins (MeCP) and complexes with histone deacetylase (HDAC) activity. Chromatin is in a condensed, nonpermissive conformation that inhibits transcription. Clinical trials are under way utilizing the combination of demethylating agents such as 5-aza-2'-deoxycytidine plus HDAC inhibitors, which together confer an open, permissive chromatin structure (*lower portion*). Transcription factors bind to specific DNA sequences in promoter regions and, through protein-protein interactions, recruit coactivator complexes containing histone acetyl transferase (HAT) activity. This enhances transcription initiation by RNA polymerase II and associated general transcription factors. The expression of the tumor-suppressor gene commences, with phenotypic changes that may include growth arrest, differentiation, or apoptosis.

This arrests differentiation at the promyelocyte stage and promotes tumor cell proliferation and survival. Treatment with pharmacologic doses of all-*trans* retinoic acid (ATRA), the ligand for RAR α , results in the release of HDAC activity and the recruitment of coactivators, which overcome the differentiation block. This induced differentiation of APL cells has improved treatment of these patients but also has led to a novel treatment toxicity when newly differentiated tumor cells infiltrate the lungs. ATRA represents a treatment paradigm for the reversal of epigenetic changes in cancer. Other leukemia-associated fusion proteins, such as Tel-acute myeloid leukemia (AML1), AML1-eight-twenty-one (ETO), and the MLL fusion proteins seen in acute myeloid leukemia (AML) and acute lymphocytic leukemia, also lead to repression through the HDAC complex. Therefore, efforts are ongoing to determine the structural basis for interactions between translocation fusion proteins and chromatin-remodeling proteins and to use this information to rationally design small molecules that will disrupt specific protein-protein associations, although this has proven to be technically difficult. Several drugs that block the enzymatic activity of HDACs (HDAC inhibitors [HDACis]) are approved for cancer treatment, and others are being tested. HDACis have demonstrated sufficient antitumor activity against cutaneous T-cell lymphoma (vorinostat, romidepsin), peripheral T-cell lymphoma (romidepsin, belinostat), and multiple myeloma (panobinostat) to be approved by the U.S. Food and Drug Administration (FDA).

HDACis have also demonstrated antitumor activity in clinical studies against some solid tumors, and additional studies are ongoing. HDACis may target cancer cells via a number of mechanisms including both epigenetic modulation via histone acetylation and effects on other proteins that are acetylated. The pleiotropic effects of some HDACis include enhancement of apoptosis by upregulation of a number of proteins that enhance apoptosis including death receptors (DR4/5, FAS, and their ligands) and downregulation of proteins that

inhibit apoptosis (e.g., X-linked inhibitor of apoptosis [XIAP]); upregulation of proteins that inhibit cell cycle progression (e.g., p21Cip1/Waf1); inhibition of DNA repair and generation of ROS leading to increased DNA damage; and disruption of the chaperone protein HSP90.

Efforts are also under way to modulate other epigenetic processes such as reversing the hypermethylation of CpG islands that characterizes many malignancies. Drugs that induce DNA demethylation, such as 5-aza-2'-deoxycytidine, can lead to reexpression of silenced genes in cancer cells with restoration of function, and 5-aza-2'-deoxycytidine is approved for use in myelodysplastic syndrome. However, 5-aza-2'-deoxycytidine has limited aqueous solubility and is myelosuppressive, limiting its usefulness. Other inhibitors of DNA methyltransferases are in development. In ongoing clinical trials, inhibitors of DNA methylation are being combined with HDACis, with the idea that reversing coexisting epigenetic changes will reverse the deregulated patterns of gene transcription in cancer cells.

Epigenetic gene regulation can also occur via microRNAs or long noncoding RNAs (lncRNA). MicroRNAs (miRNA) are short (average 22 nucleotides in length) RNA molecules that silence gene expression after transcription by binding and inhibiting the translation or promoting the degradation of mRNA transcripts. It is estimated that >1000 miRNAs are encoded in the human genome. Each tissue has a distinctive repertoire of miRNA expression, and this pattern is altered in specific ways in cancers. Specific correlations between miRNA expression and tumor biology and clinical behavior are continuing to emerge. Therapies targeting miRNAs are not currently at hand but represent an ongoing area of treatment development. lncRNAs are longer than 200 nucleotides and comprise the largest group of noncoding RNAs. Some of them have been shown to play important roles in gene regulation. The potential for altering these RNAs for therapeutic benefit is an area of active investigation.

518 APOPTOSIS AND OTHER MECHANISMS OF CELL DEATH

Tissue homeostasis requires a balance between the death of aged, terminally differentiated cells or severely damaged cells and their renewal by proliferation of committed progenitors. Genetic damage to growth-regulating genes of stem cells could lead to catastrophic results for the host as a whole. Thus, genetic events causing activation of oncogenes or loss of tumor suppressors, which would be predicted to lead to unregulated cell proliferation unless corrected, usually activate signal transduction pathways that block aberrant cell proliferation. These pathways can lead to a form of programmed cell death (*apoptosis*) or irreversible growth arrest (*senescence*). Much as a panoply of intra- and extracellular signals impinge upon the core cell cycle machinery to regulate cell division, so too these signals are transmitted to a core enzymatic machinery that regulates cell death and survival.

Apoptosis is a tightly regulated process induced by two main pathways (Fig. 72-5). The extrinsic pathway of apoptosis is activated by

cross-linking members of the tumor necrosis factor (TNF) receptor superfamily, such as CD95 (Fas) and death receptors DR4 and DR5, by their ligands, Fas ligand or TRAIL (TNF-related apoptosis-inducing ligand), respectively. This induces the association of FADD (Fas-associated death domain) and pro-caspase-8 to death domain motifs of the receptors. Caspase-8 is activated and then cleaves and activates effector caspases-3 and -7, which then target cellular constituents (including caspase-activated DNase, cytoskeletal proteins, and a number of regulatory proteins), inducing the morphologic appearance characteristic of apoptosis, which pathologists term *karyorrhexis*. The intrinsic pathway of apoptosis is initiated by the release of cytochrome c and SMAC (second mitochondrial activator of caspases) from the mitochondrial intermembrane space in response to a variety of noxious stimuli, including DNA damage, loss of adherence to the extracellular matrix (ECM), oncogene-induced proliferation, and growth factor deprivation. Upon release into the cytoplasm, cytochrome c associates with dATP, pro-caspase-9, and the adaptor protein APAF-1, leading to the

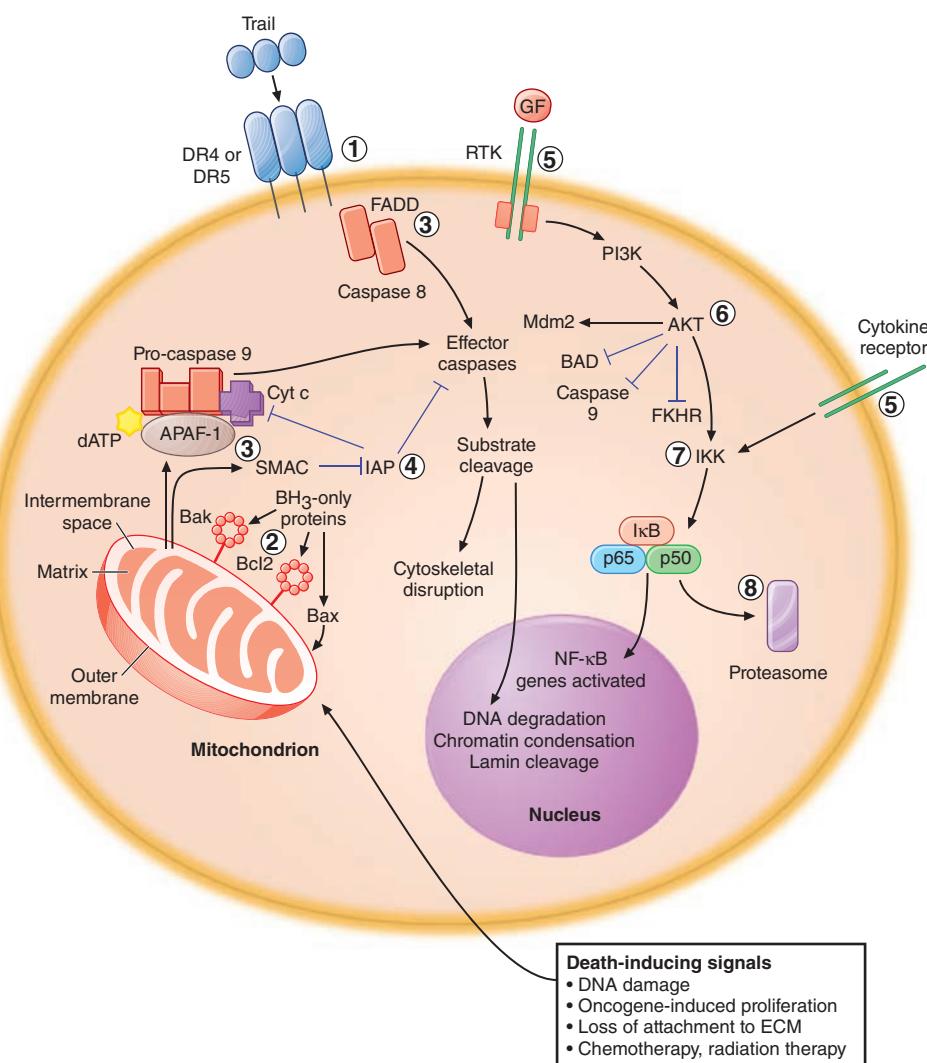


FIGURE 72-5 Therapeutic strategies to overcome aberrant survival pathways in cancer cells. 1. The extrinsic pathway of apoptosis can be selectively induced in cancer cells by TRAIL (the ligand for death receptors 4 and 5) or by agonistic monoclonal antibodies. 2. Inhibition of antiapoptotic Bcl-2 family members with antisense oligonucleotides or inhibitors of the BH₃-binding pocket will promote formation of Bak- or Bax-induced pores in the mitochondrial outer membrane. 3. Epigenetic silencing of APAF-1, caspase-8, and other proteins can be overcome using demethylating agents and inhibitors of histone deacetylases. 4. Inhibitor of apoptosis proteins (IAP) blocks activation of caspases; small-molecule inhibitors of IAP function (mimicking SMAC action) should lower the threshold for apoptosis. 5. Signal transduction pathways originating with activation of receptor tyrosine kinase receptors (RTKs) or cytokine receptors promote survival of cancer cells by a number of mechanisms. Inhibiting receptor function with monoclonal antibodies, such as trastuzumab or cetuximab, or inhibiting kinase activity with small-molecule inhibitors can block the pathway. 6. The Akt kinase phosphorylates many regulators of apoptosis to promote cell survival; inhibitors of Akt may render tumor cells more sensitive to apoptosis-inducing signals; however, the possibility of toxicity to normal cells may limit the therapeutic value of these agents. 7 and 8. Activation of the transcription factor NF-κB (composed of p65 and p50 subunits) occurs when its inhibitor, IκB, is phosphorylated by IκB-kinase (IKK), with subsequent degradation of IκB by the proteasome. Inhibition of IKK activity should selectively block the activation of NF-κB target genes, many of which promote cell survival. Inhibitors of proteasome function are U.S. Food and Drug Administration approved and may work in part by preventing destruction of IκB, thus blocking NF-κB nuclear localization. NF-κB is unlikely to be the only target for proteasome inhibitors.

sequential activation of caspase-9 and effector caspases. SMAC binds to and blocks the function of inhibitor of apoptosis proteins (IAP), negative regulators of caspase activation.

The release of apoptosis-inducing proteins from the mitochondria is regulated by pro- and antiapoptotic members of the Bcl-2 family. Antiapoptotic members (e.g., Bcl-2, Bcl-XL, and Mcl-1) associate with the mitochondrial outer membrane via their carboxyl termini, exposing to the cytoplasm a hydrophobic binding pocket composed of Bcl-2 homology (BH) domains 1, 2, and 3 that is crucial for their activity. Perturbations of normal physiologic processes in specific cellular compartments lead to the activation of BH3-only proapoptotic family members (e.g., Bad, Bim, Bid, Puma, Noxa, and others) that can alter the conformation of the outer-membrane proteins Bax and Bak, which then oligomerize to form pores in the mitochondrial outer membrane resulting in cytochrome c release. If proteins comprised only by BH3 domains are sequestered by Bcl-2, Bcl-XL, or Mcl-1, pores do not form and apoptosis-inducing proteins are not released from the mitochondria. The ratio of levels of antiapoptotic Bcl-2 family members and the levels of proapoptotic BH3-only proteins at the mitochondrial membrane determines the activation state of the intrinsic pathway. The mitochondrion must therefore be recognized not only as an organelle with vital roles in intermediary metabolism and oxidative phosphorylation but also as a central regulatory structure of the apoptotic process.

The evolution of tumor cells to a more malignant phenotype requires the acquisition of genetic changes that subvert apoptosis pathways and promote cancer cell survival and resistance to anticancer therapies. However, cancer cells may be more vulnerable than normal cells to therapeutic interventions that target the apoptosis pathways that cancer cells depend upon. For instance, overexpression of Bcl-2 as a result of the t(14;18) translocation contributes to follicular lymphoma, and it is highly expressed in many lymphoid malignancies including chronic lymphocytic leukemia (CLL). Upregulation of Bcl-2 expression is also observed in other cancers including prostate, breast, and lung cancers and melanoma. Targeting of antiapoptotic Bcl-2 family members has been accomplished by the identification of several low-molecular-weight compounds that bind to the hydrophobic pockets of either Bcl-2 or Bcl-XL and block their ability to associate with death-inducing BH3-only proteins. An oral BH3 mimetic inhibitor of BCL-2, venetoclax, is approved for use in patients with refractory CLL with 17p deletion and is active in acute myeloid leukemia.

Preclinical studies targeting death receptors DR4 and -5 have demonstrated that recombinant, soluble, human TRAIL or humanized monoclonal antibodies with agonist activity against DR4 or -5 can induce apoptosis of tumor cells while sparing normal cells. The mechanisms for this selectivity may include expression of decoy receptors or elevated levels of intracellular inhibitors (such as FLIP, which competes with caspase-8 for FADD) by normal cells but not tumor cells. Synergy has been shown between TRAIL-induced apoptosis and chemotherapeutic agents in some preclinical studies. However, studies have not yet shown significant clinical activity of approaches targeting the TRAIL pathway.

Many of the signal transduction pathways perturbed in cancer promote tumor cell survival (Fig. 72-5). These include activation of the PI3K/Akt pathway, increased levels of the NF- κ B transcription factor, and epigenetic silencing of genes such as APAF-1 (apoptosis protease activating factor-1 involved in activating caspase-9 and essential for apoptosis) and caspase-8. Each of these pathways is a target for therapeutic agents that, in addition to affecting cancer cell proliferation or gene expression, may render cancer cells more susceptible to apoptosis, thus promoting synergy when combined with other chemotherapeutic agents.

Some tumor cells resist drug-induced apoptosis indirectly by eliminating the noxious stimulus-inducing apoptosis through expression of one or more members of the ABC (ATP-binding cassette proteins) family of ATP-dependent efflux pumps that mediate the multidrug-resistance (MDR) phenotype. The prototype member of this family, P-glycoprotein (PGP), spans the plasma membrane 12 times and has two ATP-binding sites. Hydrophobic drugs (e.g., anthracyclines and vinca alkaloids) are recognized by PGP as they enter the cell and are

pumped out. Numerous clinical studies have failed to demonstrate that drug resistance can be overcome using inhibitors of PGP. However, ABC transporters have different substrate specificities, and inhibition of a single family member may not be sufficient to overcome the MDR phenotype. Efforts to reverse PGP-mediated drug resistance continue.

Cells, including cancer cells, can also undergo other mechanisms of cell death including *autophagy* (degradation of proteins and organelles by lysosomal proteases) and *necrosis* (digestion of cellular components and rupturing of the cell membrane). Necrosis usually occurs in response to external forces resulting in release of cellular components, which leads to inflammation and damage to surrounding tissues. Although necrosis was thought to be unprogrammed, evidence now suggests that at least some aspects may also be programmed. The exact role of necrosis in cancer cell death in various settings is still being determined. In addition to its role in cell death, autophagy can also serve as a homeostatic mechanism to promote survival for the cell by recycling cellular components to provide necessary energy. The mechanisms that control the balance between enhancing survival versus leading to cell death are still not fully understood. Autophagy appears to play conflicting roles in the development and survival of cancer. Early in the carcinogenic process, it can act as a tumor suppressor by preventing the cell from accumulating abnormal proteins and organelles. However, in established tumors, it may serve as a mechanism of survival for cancer cells when they are stressed by damage such as from chemotherapy. Preclinical studies have indicated that inhibition of this process can enhance the sensitivity of cancer cells to chemotherapy or radiation therapy, and ongoing trials are evaluating inhibitors of autophagy in combination with chemotherapy and/or radiation therapy. Better understanding of the factors that control the survival-promoting versus death-inducing aspects of autophagy is required in order to know how to best manipulate it for therapeutic benefit.

METASTASIS

The metastatic process accounts for the vast majority of deaths from solid tumors, and therefore, an understanding of this process is critical for improvements in survival from cancer. The biology of metastasis is complex and requires multiple steps. The initial step involves cell migration and invasion through the ECM. The three major features of tissue invasion are cell adhesion to the basement membrane, local proteolysis of the membrane, and movement of the cell through the rent in the membrane and the ECM. Cells that lose contact with the ECM normally undergo programmed cell death (anoikis-apoptosis induced by the loss of contact), and this process has to be suppressed in cells that metastasize. Another process important for many, but not necessarily all, metastasizing epithelial cancer cells is epithelial-mesenchymal transition (EMT). This is a process by which cells lose their epithelial properties and gain mesenchymal properties. This normally occurs during the developmental process in embryos, allowing cells to migrate to their appropriate destinations in the embryo. It also occurs in wound healing, tissue regeneration, and fibrotic reactions, but in all of these processes, cells stop proliferating when the process is complete. Malignant cells that metastasize often undergo EMT as an important step in that process but retain the capacity for unregulated proliferation. However, there is evidence that not all metastasizing cancer cells require EMT, and the exact role of EMT in different metastasizing cancer cells continues to be elucidated. Malignant cells that gain access to the circulation must then repeat those steps at a remote site, find a hospitable niche in a foreign tissue, avoid detection and elimination by host defenses including the immune system, and induce the growth of new blood vessels. Some metastatic cells occur as oligoclonal clusters, which appear to be more potent in establishing metastasis than single cells, perhaps, in part, through differential and cooperative effects in evading host defenses. The rate-limiting step for metastasis is the ability for tumor cells to survive and expand in the novel microenvironment of the metastatic site, and multiple host-tumor interactions determine the ultimate outcome (Fig. 72-6). Few drugs have been developed to attempt to directly target the process of metastasis, in part because the specifics of the critical steps in the process that would be potentially good targets for drugs are still being

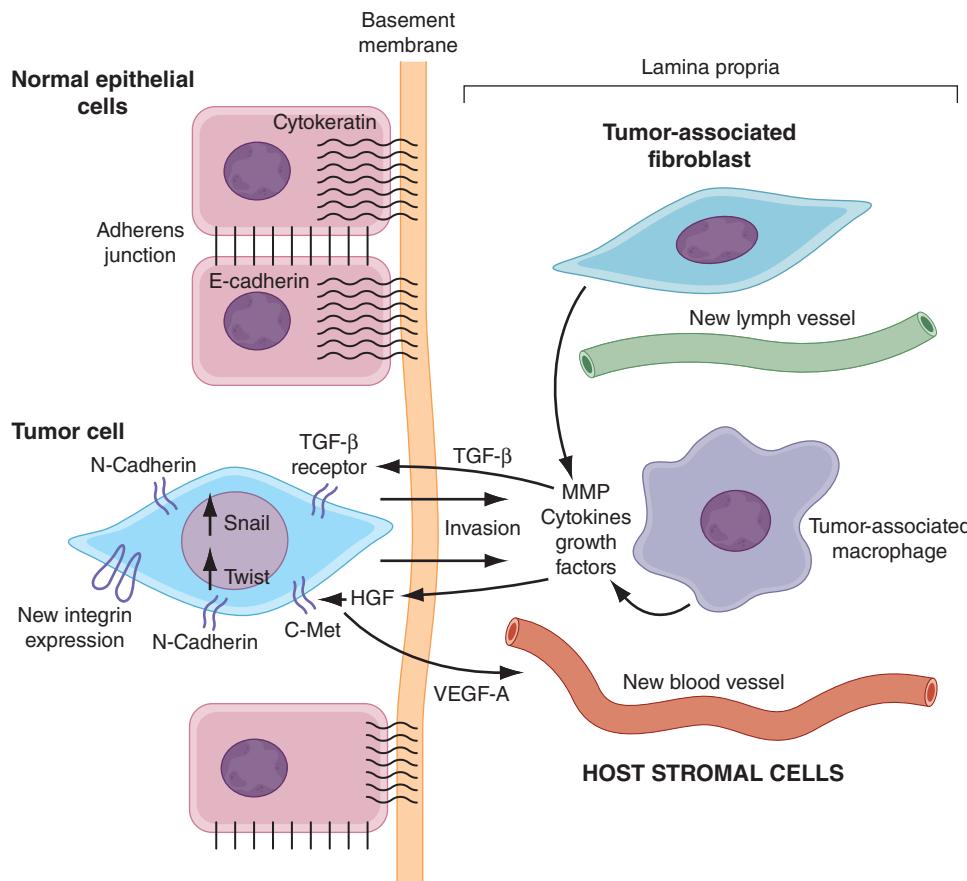


FIGURE 72-6 Oncogene signaling pathways are activated during tumor progression and promote metastatic potential. This figure shows a cancer cell that has undergone epithelial to mesenchymal transition (EMT) under the influence of several environmental signals. Critical components include activated transforming growth factor beta (TGF- β) and the hepatocyte growth factor (HGF)/c-Met pathways, as well as changes in the expression of adhesion molecules that mediate cell-cell and cell-extracellular matrix interactions. Important changes in gene expression are mediated by the Snail and Twist family of transcriptional repressors (whose expression is induced by the oncogenic pathways), leading to reduced expression of E-cadherin, a key component of adherens junctions between epithelial cells. This, in conjunction with upregulation of N-cadherin, a change in the pattern of expression of integrins (which mediate cell-extracellular matrix associations that are important for cell motility), and a switch in intermediate filament expression from cytokeratin to vimentin, results in the phenotypic change from adherent highly organized epithelial cells to motile and invasive cells with a fibroblast or mesenchymal morphology. EMT is thought to be an important step leading to metastasis in some human cancers. Host stromal cells, including tumor-associated fibroblasts and macrophages, play an important role in modulating tumor cell behavior through secretion of growth factors and proangiogenic cytokines, and matrix metalloproteinases that degrade the basement membrane. VEGF-A, -C, and -D are produced by tumor cells and stromal cells in response to hypoxemia or oncogenic signals and induce production of new blood vessels and lymphatic channels through which tumor cells metastasize to lymph nodes or tissues.

identified. However, a number of potential targets are known. HER2 can enhance the metastatic potential of breast cancer cells, and as discussed above, the monoclonal antibody trastuzumab, which targets HER2, improves survival in the adjuvant setting for HER2-positive breast cancer patients. A number of other potential targets that increase metastatic potential of cells in preclinical studies include HIF-1 and -2, transcription factors induced by hypoxia within tumors, growth factors (e.g., cMET and VEGFR), oncogenes (e.g., SRC), adhesion molecules (e.g., focal adhesion kinase [FAK]), ECM proteins (e.g., matrix metalloproteinases 1 and 2), and inflammatory molecules (e.g., COX-2).

The metastatic phenotype is likely restricted to a fraction of tumor cells (Fig. 72-6). A number of genetic and epigenetic changes are required for tumor cells to be able to metastasize, including activation of metastatic-promoting genes and inhibition of genes that suppress the metastatic ability. Given the role of microRNAs in controlling gene expression (see epigenetic section) including those critical to the metastatic process, efforts are under way to modulate these to try to inhibit metastasis. Cells with metastatic capability frequently express chemokine receptors that are likely important in the metastatic process. A number of candidate metastasis-suppressor genes have been identified, including genes coding for proteins that enhance apoptosis, suppress cell division, are involved in the interactions of cells with each other or the ECM, or suppress cell migration. The loss of function of these genes enhances metastasis. Gene expression profiling is being used to

study the metastatic process and other properties of tumor cells that may predict susceptibilities.

An example of the ability of malignant cells to survive and grow in a novel microenvironment is bone metastases. Bone metastases can be extremely painful, cause fractures of weight-bearing bones, can lead to hypercalcemia, and are a major cause of morbidity for cancer patients. Osteoclasts and their monocyte-derived precursors express the surface receptor RANK (receptor activator of NF- κ B), which is required for terminal differentiation and activation of osteoclasts. Osteoblasts and other stromal cells express RANK ligand (RANKL), as both a membrane-bound and soluble cytokine. Osteoprotegerin (OPG), a soluble receptor for RANKL produced by stromal cells, acts as a decoy receptor to inhibit RANK activation. The relative balance of RANKL and OPG determines the activation state of RANK on osteoclasts. Many tumors increase osteoclast activity by secretion of substances such as parathyroid hormone (PTH), PTH-related peptide, interleukin (IL) 1, or Mip1 that perturb the homeostatic balance of bone remodeling by increasing RANK signaling. One example is multiple myeloma, where tumor cell-stromal cell interactions activate osteoclasts and inhibit osteoblasts, leading to the development of multiple lytic bone lesions. Inhibition of RANKL by an antibody (denosumab) can prevent further bone destruction. Bisphosphonates are also effective inhibitors of osteoclast function that are used in the treatment of cancer patients with bone metastases.

CANCER STEM CELLS

Normal tissues have stem cells capable of self-renewal and repairing damaged tissue, whereas the majority of cells in normal tissues do not have this capacity. Similarly, only a small proportion of the cells within a tumor are capable of initiating colonies *in vitro* or forming tumors at high efficiency when injected into immunocompromised NOD/SCID mice. For example, AML and CML have a small population of cells (estimated to be <1%) that have properties of stem cells, such as unlimited self-renewal and the capacity to cause leukemia when serially transplanted in mice. These cells have an undifferentiated phenotype (Thy1-CD34+CD38- and do not express other differentiation markers) and resemble normal stem cells in many ways but are no longer under homeostatic control (Fig. 72-7). Solid tumors may also contain a population of stem cells. It is not yet known how often cancers may originate within a stem cell population. Cancer stem cells, like their normal counterparts, have unlimited proliferative capacity and paradoxically traverse the cell cycle at a slow rate; cancer growth occurs largely due to expansion of the stem cell pool, the unregulated proliferation of an amplifying population, and failure of apoptosis pathways (Fig. 72-7). Slow cell cycle progression and high levels of expression of antiapoptotic Bcl-2 family members and drug efflux pumps of the MDR family render cancer stem cells less vulnerable to cancer chemotherapy or radiation therapy. Implicit in the cancer stem cell hypothesis is the idea that failure to cure most human cancers is due to the fact that current therapeutic agents do not kill the stem cells. Identification and isolation of cancer stem cells will allow determination of the aberrant signaling pathways that distinguish these cells from

normal tissue stem cells. These would serve as potential therapeutic targets. Evidence that cells with stem cell properties can arise from other epithelial cells within the cancer by processes such as epithelial-mesenchymal transition also implies that it is essential to treat all of the cancer cells, and not just those with current stem cell-like properties, in order to eliminate the self-renewing cancer cell population. The exact nature of cancer stem cells remains an area of investigation. One of the unanswered questions is the exact origin of cancer stem cells for the different cancers.

PLASTICITY AND RESISTANCE

Cancer cells, and especially stem cells, have the capacity for significant plasticity allowing them to alter multiple aspects of cell biology in response to external factors (e.g., chemotherapy, radiation therapy, inflammation, immune response). In addition, heterogeneity between the different clones of cells within the tumor population and their interactions with each other and the tumor microenvironment provides the tumor with the capacity for significant plasticity in dealing with both internal and external stresses. Thus, a major problem in cancer therapy is that malignancies have a wide spectrum of mechanisms for both initial and adaptive resistance to treatments. These include inhibiting drug delivery to the cancer cells, blocking drug uptake and retention, increasing drug metabolism, altering levels of target proteins making them less sensitive to drugs, acquiring mutations in target proteins making them no longer sensitive to the drug, modifying metabolism and cell signaling pathways, using alternate signaling pathways, adjusting the cell replication process including mechanisms

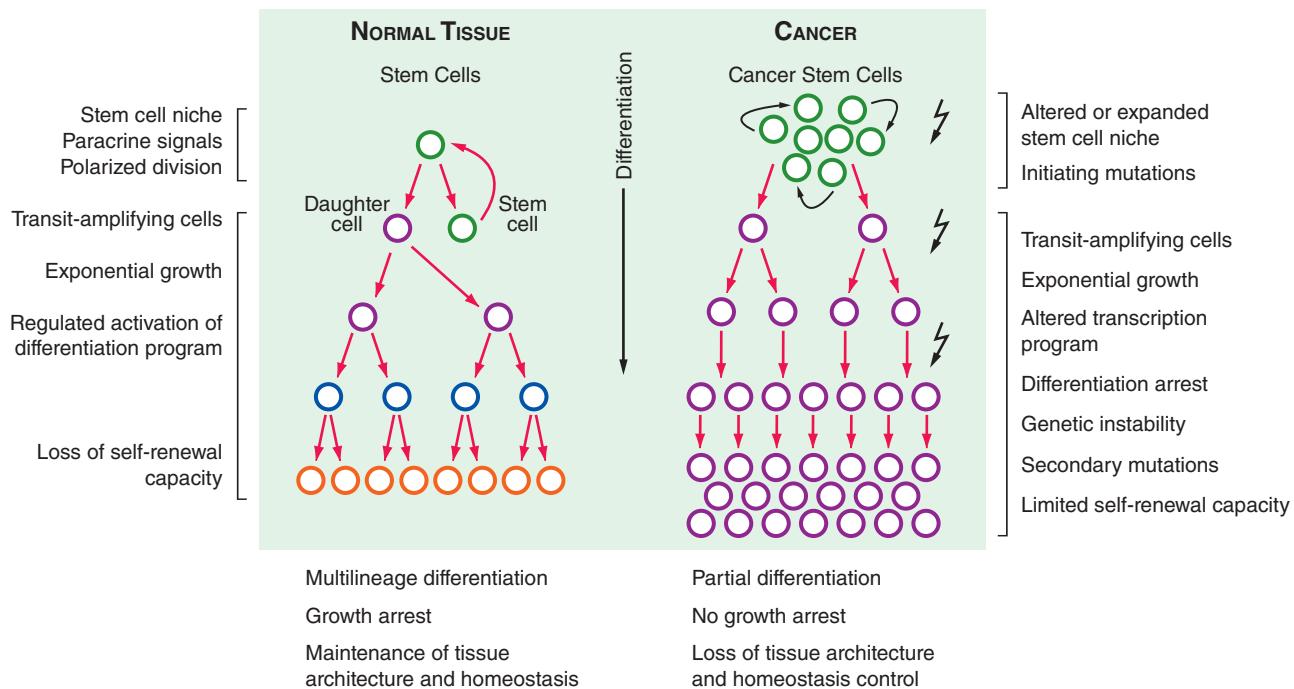


FIGURE 72-7 **Cancer stem cells play a critical role in the initiation, progression, and resistance to therapy of malignant neoplasms.** In normal tissues (*left*), homeostasis is maintained by asymmetric division of stem cells leading to one progeny cell that will differentiate and one cell that will maintain the stem cell pool. This occurs within highly specific niches unique to each tissue, such as in close apposition to osteoblasts in bone marrow, or at the base of crypts in the colon. Here, paracrine signals from stromal cells, such as sonic hedgehog or Notch ligands, as well as upregulation of β -catenin and telomerase, help to maintain stem cell features of unlimited self-renewal while preventing differentiation or cell death. This occurs in part through upregulation of the transcriptional repressor Bmi-1 and inhibition of the p16^{Ink4a}/Arf and p53 pathways. Daughter cells leave the stem cell niche and enter a proliferative phase (referred to as *transit-amplifying*) for a specified number of cell divisions, during which time a developmental program is activated, eventually giving rise to fully differentiated cells that have lost proliferative potential. Cell renewal equals cell death, and homeostasis is maintained. In this hierarchical system, only stem cells are long-lived. The hypothesis is that cancers harbor stem cells that make up a small fraction (i.e., 0.001–1%) of all cancer cells. These cells share several features with normal stem cells, including an undifferentiated phenotype, unlimited self-renewal potential, and a capacity for some degree of differentiation; however, due to initiating mutations (mutations are indicated by lightning bolts), they are no longer regulated by environmental cues. The cancer stem cell pool is expanded, and rapidly proliferating progeny, through additional mutations, may attain stem cell properties, although most of this population is thought to have a limited proliferative capacity. Differentiation programs are dysfunctional due to reprogramming of the pattern of gene transcription by oncogenic signaling pathways. Within the cancer transit-amplifying population, genomic instability generates aneuploidy and clonal heterogeneity as cells attain a fully malignant phenotype with metastatic potential. The cancer stem cell hypothesis has led to the idea that current cancer therapies may be effective at killing the bulk of tumor cells but do not kill tumor stem cells, leading to a regrowth of tumors that is manifested as tumor recurrence or disease progression. Research is in progress to identify unique molecular features of cancer stem cells that can lead to their direct targeting by novel therapeutic agents.

by which the cell deals with DNA damage, inhibiting apoptosis, and evading the immune system. Thus, most metastatic cancers (except those curable with chemotherapy such as germ cell tumors) eventually become resistant to the therapy being utilized. Overcoming resistance is a major area of research.

CANCER METABOLISM

One of the distinguishing characteristics of cancer cells is that they have altered metabolism as compared with normal cells in supporting survival, their high rates of proliferation, and ability to metastasize. Complicating studies evaluating metabolic differences between normal and malignant cells is that there is heterogeneity in metabolism between different cells within a cancer. Malignant cells must focus a significant fraction of their energy resources into synthesis of proteins and other molecules (building blocks required for the production of new cells) while still maintaining sufficient ATP production to survive and grow. Although normal proliferating cells also have similar needs, there are differences in how cancer cells metabolize glucose and a number of other compounds including the amino acid glutamine as compared to normal cells in part because of genetic and epigenetic changes within cancer cells but also likely due to differences in the environments of cancer and normal cells. Many cancer cells utilize aerobic glycolysis (the Warburg effect) (Fig. 72-8) to metabolize glucose, leading to increased lactic acid production, whereas normal cells utilize oxidative phosphorylation in mitochondria under aerobic conditions, a much more efficient process for generating ATP for energy utilization but one that does not produce the same level of building blocks needed for new cells. One consequence is increased glucose uptake and utilization by cancer cells, a fact utilized in fluorodeoxyglucose (FDG)-positron emission tomography (PET) scanning to detect tumors. A number of proteins in cancer cells, including cMYC, HIF1, RAS, p53, pRB, and AKT, are involved in modulating glycolytic processes and controlling the Warburg effect. Although these pathways remain difficult to target therapeutically, both the PI3K pathway with signaling through mTOR and the AMP-activated kinase (AMPK) pathway that inhibits mTORC1 (a protein complex that includes mTOR) are important in controlling the glycolytic process and thus provide potential targets for inhibiting this process. An inhibitor of mTOR is approved for use against RCC (temsirolimus), and another inhibitor (everolimus) has activity against breast and neuroendocrine cancer and RCC. Other mTOR inhibitors

are in trials, and modulators of AMPK are being investigated. The inefficient utilization of glucose by malignant cells also leads to a need for alternative metabolic pathways for other compounds as well, one of which is glutamine. Similar to glucose, this provides both a source for structural molecules as well as energy production. Similarly to glucose, glutamine is also inefficiently utilized by cancer cells. Cancer cells can also take up nutrients released by surrounding cells and tissues, increasing the complexity of successfully therapeutically inhibiting metabolism in cancer.

Mutations in genes involved in the metabolic process occur in a number of cancers. Among the most frequently found to date are mutations in isocitrate dehydrogenases 1 and 2 (IDH1 and IDH2). These have been most commonly seen in gliomas, AMLs, and intrahepatic cholangiocarcinomas. These mutations lead to the production of an oncometabolite (2-hydroxyglutarate [2HG]) instead of the normal product α -ketoglutarate. Although the exact mechanisms of oncogenesis by 2HG are still being elucidated, α -ketoglutarate is a key cofactor for a number of dioxygenases involved in controlling DNA methylation. 2HG can act as a competitive inhibitor for α -ketoglutarate, leading to alterations in methylation status (primarily hypermethylation) of genes (leading to epigenetic changes) that can have profound effects on a number of cellular processes including differentiation. Inhibitors of mutant IDH1 and IDH2 are approved for treating IDH mutant AML and are in clinical trials for glioblastomas and cholangiocarcinomas.

Much needs to be learned about the specific differences in metabolism between cancer cells and normal cells; however, even with the currently limited state of knowledge, modulators of metabolism are being tested clinically. The first of these is the antidiabetic agent metformin, both alone and in combination with chemotherapeutic agents. Metformin inhibits gluconeogenesis and may have direct effects on tumor cells by activating AMPK, a serine/threonine protein kinase that is downstream of the LKB1 tumor suppressor, and thus inhibiting mTOR complex 1 (mTORC1). This leads to decreased protein synthesis and proliferation. Studies to date have not yet established metformin to have a clear role as an anticancer agent.

TUMOR MICROENVIRONMENT, ANGIOGENESIS, AND IMMUNE EVASION

Tumors consist not only of malignant cells but also of a complex microenvironment including many other types of cells (including lymphocytes,

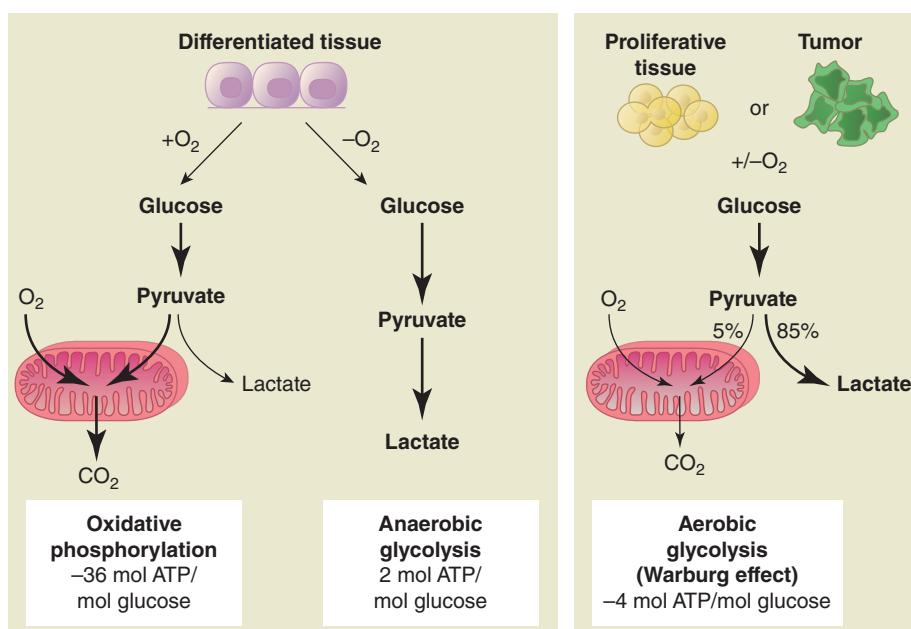


FIGURE 72-8 Warburg versus oxidative phosphorylation. In most normal tissues, the vast majority of cells are differentiated and dedicated to a particular function within the organ in which they reside. The metabolic needs are mainly for energy and not for building blocks for new cells. In these tissues, ATP is generated by oxidative phosphorylation that efficiently generates about 36 molecules of ATP for each molecule of glucose metabolized. By contrast, proliferative tumor tissues, especially in the setting of hypoxia, a typical condition within tumors, use aerobic glycolysis to generate energy for cell survival and generation of building blocks for new cells.

macrophages, myeloid cells, other inflammatory cells, fibroblasts, and fat cells), ECM, secreted factors (including growth factors and hormones), reactive oxygen and nitrogen species, mechanical factors, blood vessels, and lymphatics. This microenvironment is not static but rather is dynamic and continually evolving. Both the complexity and dynamic nature of the microenvironment enhance the difficulty of treating tumors. The microenvironment can contribute to resistance to anticancer therapies through a number of mechanisms.

OBESITY AND CANCER

Significant evidence links obesity and the increased risk of developing certain cancers including postmenopausal breast, colorectal, ovarian, endometrial, esophageal, gallbladder, thyroid, and kidney cancers, among others. Less certain are the mechanisms responsible for this risk. As outlined above, cancers arise in an environment with multiple factors, many of which can stimulate cell proliferation. Obesity impacts a variety of factors including hormonal factors, altered metabolism (especially adipose metabolism), and mediators of inflammatory response that all can impact the development of malignancy. Obesity is associated with a number of hormonal changes including high insulin, glucagon, and leptin levels that can stimulate growth of cells. It also leads to insulin resistance, which may contribute to cancer cell development, in part by increasing insulin-like growth factor-1 (IGF-1) levels. Obesity also leads to alterations in adipose, including fatty acid, metabolism, with production of compounds important for energy metabolism as well as for membrane function within cells that may contribute to carcinogenic process. Obesity contributes to an inflammatory environment in a variety of ways including increased levels of inflammatory proteins such as IL-6 and TNF- α . In terms of impact on survival with cancer, data primarily from breast cancer suggest that obesity is associated with decreased survival likely due, at least in part, to the impact of obesity on hormonal factors in development of certain breast cancers, although this may be limited to subsets of breast cancer patients. Some studies have suggested, paradoxically, that obesity may be associated with improved survival in some patients such as those with advanced-stage colorectal cancer. Clearly, the biology of the association between obesity and cancer and its impact on disease outcome is complex, and additional studies are necessary to better define the mechanisms involved.

MECHANISMS OF TUMOR VESSEL FORMATION

One of the critical elements of tumor cell proliferation is delivery of oxygen, nutrients, and circulating factors important for growth and survival. Thus, a critical element in growth of primary tumors and formation of metastatic sites is the *angiogenic switch*: the ability of the tumor to promote the formation of new blood vessels, including the recruitment of vascular endothelial cells (ECs). The angiogenic switch is a phase in tumor development when the dynamic balance of pro- and antiangiogenic factors is tipped in favor of vessel formation by the effects of the tumor on its immediate environment. Stimuli for tumor angiogenesis include hypoxemia, inflammation, and genetic lesions in oncogenes or tumor suppressors that alter tumor cell gene expression. Angiogenesis consists of several steps, including the stimulation of ECs by growth factors, degradation of the ECM by proteases, proliferation and migration of ECs into the tumor, and the eventual formation of new capillary tubes.

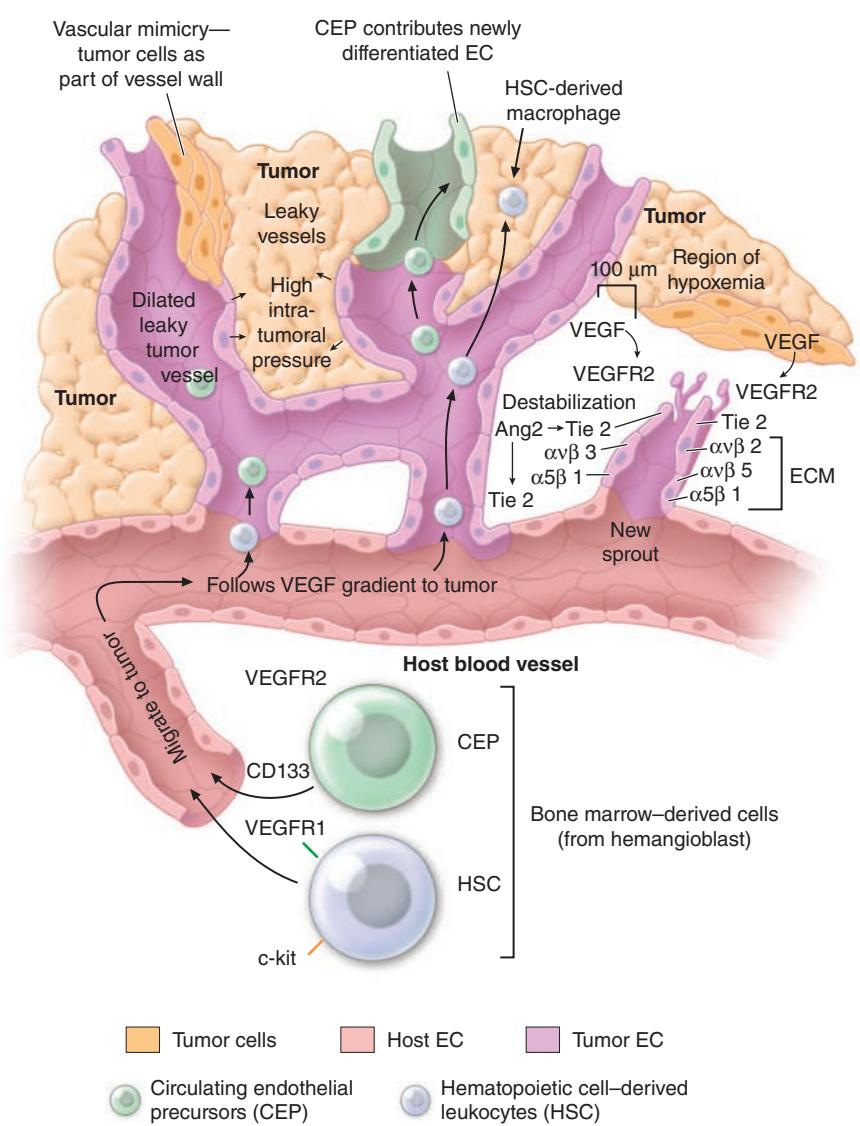


FIGURE 72-9 Tumor angiogenesis is a complex process involving many different cell types that must proliferate, migrate, invade, and differentiate in response to signals from the tumor microenvironment. Endothelial cells (ECs) sprout from host vessels in response to VEGF, bFGF, Ang2, and other proangiogenic stimuli. Sprouting is stimulated by VEGF/VEGFR2, Ang2/Tie-2, and integrin/extracellular matrix (ECM) interactions. Bone marrow-derived circulating endothelial precursors (CEPs) migrate to the tumor in response to VEGF and differentiate into ECs, while hematopoietic stem cells differentiate into leukocytes, including tumor-associated macrophages that secrete angiogenic growth factors and produce matrix metalloproteinases (MMPs) that remodel the ECM and release bound growth factors. Tumor cells themselves may directly form parts of vascular channels within tumors. The pattern of vessel formation is haphazard: vessels are tortuous, dilated, leaky, and branch in random ways. This leads to uneven blood flow within the tumor, with areas of acidosis and hypoxemia (which stimulate release of angiogenic factors) and high intratumoral pressures that inhibit delivery of therapeutic agents.

Tumors use a number of mechanisms to promote vascularization, subverting normal angiogenic processes for this purpose (Fig. 72-9). Primary or metastatic tumor cells sometimes arise in proximity to host blood vessels and grow around these vessels, parasitizing nutrients by co-opting the local blood supply. However, most tumor blood vessels arise by the process of sprouting, in which tumors secrete trophic angiogenic molecules, the most potent being VEGFs, that induce the proliferation and migration of host ECs into the tumor. Sprouting in normal and pathogenic angiogenesis is regulated by three families of transmembrane RTKs expressed on ECs and their ligands (VEGFs, angiopoietins, ephrins; Fig. 72-10), which are produced by tumor cells, inflammatory cells, or stromal cells in the tumor microenvironment.

Central to the angiogenic response are hypoxia-inducible factors (HIFs; especially 1 and 2), which are transcription factors that normally, in response to hypoxia, stimulate the transcription of a large number of genes responsive to hypoxia, including genes involved in metabolism as

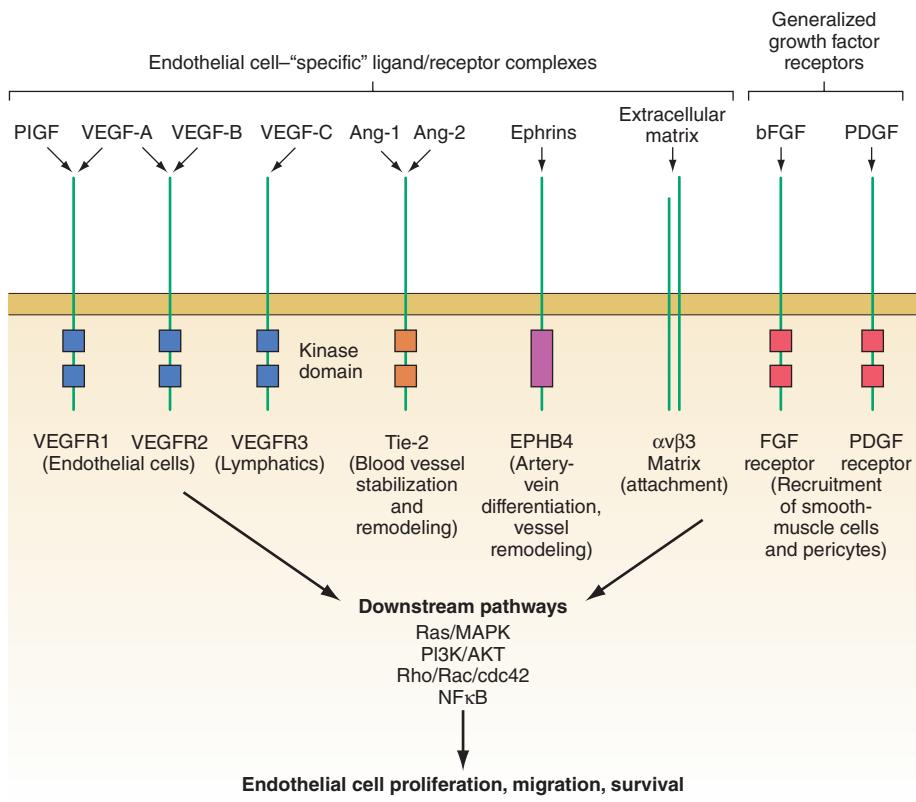


FIGURE 72-10 Critical molecular determinants of endothelial cell biology. Angiogenic endothelium expresses a number of receptors not found on resting endothelium. These include receptor tyrosine kinases (RTKs) and integrins that bind to the extracellular matrix and mediate endothelial cell (EC) adhesion, migration, and invasion. ECs also express RTKs (i.e., the fibroblast growth factor [FGF] and platelet-derived growth factor [PDGF] receptors) that are found on many other cell types. Critical functions mediated by activated RTK include proliferation, migration, and enhanced survival of endothelial cells, as well as regulation of the recruitment of perivascular cells and bloodborne circulating endothelial precursors and hematopoietic stem cells to the tumor. Intracellular signaling via EC-specific RTK utilizes molecular pathways that may be targets for future antiangiogenic therapies.

well as angiogenesis. HIF1 has a bigger role in stimulating metabolism (glycogenesis), whereas HIF2 plays a bigger role in angiogenesis. HIF protein function can also be enhanced in a number of ways in cancer not involving hypoxia, including mutations in the von Hippel-Lindau tumor suppressor gene (an E3 ubiquitin ligase that controls HIF levels by targeting it for degradation), such as occurs in some RCCs. Among the genes stimulated by HIF are VEGF and VEGF receptors. VEGFs and their receptors are required for embryonic vasculogenesis (development of new blood vessels when none preexist) and normal (wound healing, corpus luteum formation) and pathologic angiogenesis (tumor angiogenesis, inflammatory conditions such as rheumatoid arthritis). VEGF-A is a heparin-binding glycoprotein with at least four isoforms (splice variants) that regulates blood vessel formation by binding to the RTKs VEGFR1 and VEGFR2, which are expressed on all ECs in addition to a subset of hematopoietic cells (Fig. 72-9). VEGFR2 plays a more direct role in regulating EC proliferation, migration, and survival, whereas VEGFR1 appears to have more nuanced functions with a less direct role in stimulating EC processes in the normal adult (even acting as a decoy protein for VEGFA to decrease binding to VEGFR2) but with important effects during embryogenesis and on tumor angiogenesis. Tumor vessels may be more dependent on VEGFR signaling for growth and survival than normal ECs.

While VEGF signaling is a critical initiator of angiogenesis, this is a complex process regulated by additional signaling pathways (Fig. 72-10). The angiopoietin, Ang1, produced by stromal cells, binds to the EC RTK Tie-2 and promotes the interaction of ECs with the ECM and perivascular cells, such as pericytes and smooth-muscle cells, to form tight, nonleaky vessels. PDGF and basic fibroblast growth factor (bFGF) help to recruit these perivascular cells. Ang1 is required for maintaining the quiescence and stability of mature blood vessels

and prevents the vascular permeability normally induced by VEGF and inflammatory cytokines.

For tumor cell-derived VEGF to initiate sprouting from host vessels, the stability conferred by the Ang1/Tie2 pathway must be perturbed; this occurs by the secretion of Ang2 by ECs that are undergoing active remodeling. Ang2 binds to Tie2 and is a competitive inhibitor of Ang1 action: under the influence of Ang2, preexisting blood vessels become more responsive to remodeling signals, with less adherence of ECs to stroma and associated perivascular cells and more responsiveness to VEGF. Therefore, Ang2 is required at early stages of tumor angiogenesis for destabilizing the vasculature by making host ECs more sensitive to angiogenic signals. In the presence of Ang2, there is no stabilization by the Ang1/Tie2 interaction, and tumor blood vessels are leaky, hemorrhagic, and have poor association of ECs with underlying stroma. Sprouting tumor ECs express high levels of the transmembrane protein ephrin-B2 and its receptor, the RTK EPH, whose signaling appears to work with the angiopoietins during vessel remodeling. During embryogenesis, EPH receptors are expressed on the endothelium of primordial venous vessels while the transmembrane ligand ephrin-B2 is expressed by cells of primordial arteries; the reciprocal expression may regulate differentiation and patterning of the vasculature.

A number of additional ubiquitously expressed host molecules play critical roles in normal and pathologic angiogenesis. Proangiogenic cytokines, chemokines, and growth factors secreted by stromal cells or inflammatory cells make important contributions

to neovascularization, including bFGF, transforming growth factor- β (TGF- β), TNF- α , and IL-8. In contrast to normal endothelium, angiogenic endothelium overexpresses specific members of the integrin family of ECM-binding proteins that mediate EC adhesion, migration, and survival. Specifically, expression of integrins $\alpha v \beta 3$, $\alpha v \beta 5$, and $\alpha 5 \beta 1$ mediates spreading and migration of ECs and is required for angiogenesis induced by VEGF and bFGF, which in turn can upregulate EC integrin expression. The $\alpha v \beta 3$ integrin physically associates with VEGFR2 in the plasma membrane and promotes signal transduction from each receptor to promote EC proliferation (via focal adhesion kinase, src, PI3K, and other pathways) and survival (by inhibition of p53 and increasing the Bcl-2/Bax expression ratio). In addition, $\alpha v \beta 3$ forms cell-surface complexes with matrix metalloproteinases (MMPs), zinc-requiring proteases that cleave ECM proteins, leading to enhanced EC migration and the release of heparin-binding growth factors, including VEGF and bFGF. EC adhesion molecules can be upregulated (i.e., by VEGF, TNF- α) or downregulated (by TGF- β); this, together with chaotic blood flow, explains poor leukocyte-endothelial interactions in tumor blood vessels and may help tumor cells avoid immune surveillance.

Tumor blood vessels are not normal; they have chaotic architecture and blood flow. Due to an imbalance of angiogenic regulators such as VEGFs and angiopoietins (see below), tumor vessels are tortuous and dilated with an uneven diameter, excessive branching, and shunting. Tumor blood flow is variable, with areas of hypoxemia and acidosis leading to the selection of variants that are resistant to hypoxemia-induced apoptosis (often involving the loss of p53 expression). Tumor vessel walls have numerous openings, widened interendothelial junctions, and discontinuous or absent basement membrane. This contributes to the high permeability of these vessels and, together with lack of functional intratumoral lymphatics, causes increased interstitial

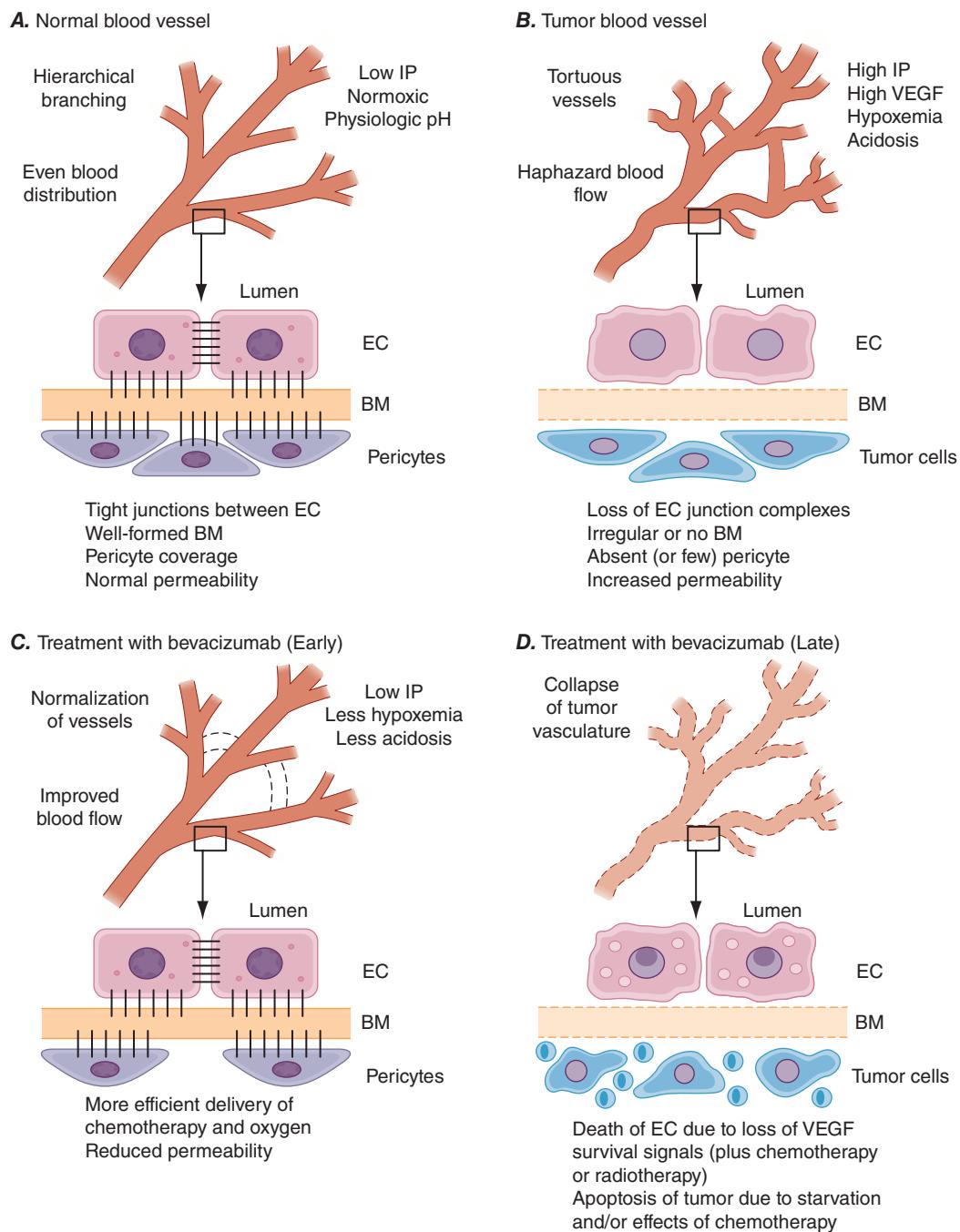


FIGURE 72-11 Normalization of tumor blood vessels due to inhibition of VEGF signaling. **A.** Blood vessels in normal tissues exhibit a regular hierarchical branching pattern that delivers blood to tissues in a spatially and temporally efficient manner to meet the metabolic needs of the tissue (top). At the microscopic level, tight junctions are maintained between endothelial cells (ECs), which are adherent to a thick and evenly distributed basement membrane (BM). Pericytes form a surrounding layer that provides trophic signals to the EC and helps maintain proper vessel tone. Vascular permeability is regulated, interstitial fluid pressure (IP) is low, and oxygen tension and pH are physiologic. **B.** Tumors have abnormal vessels with tortuous branching and dilated, irregular interconnecting branches, causing uneven blood flow with areas of hypoxemia and acidosis. This harsh environment selects genetic events that result in resistant tumor variants, such as the loss of p53. High levels of VEGF (secreted by tumor cells) disrupt gap junction communication, tight junctions, and adherens junctions between EC via src-mediated phosphorylation of proteins such as connexin 43, zonula occludens-1, VE-cadherin, and α/β -catenins. Tumor vessels have thin, irregular BM, and pericytes are sparse or absent. Together, these molecular abnormalities result in a vasculature that is permeable to serum macromolecules, leading to high tumor interstitial pressure, which can prevent the delivery of drugs to the tumor cells. This is made worse by the binding and activation of platelets at sites of exposed BM, with release of stored VEGF and microvessel clot formation, creating more abnormal blood flow and regions of hypoxemia. **C.** In experimental systems, treatment with bevacizumab or blocking antibodies to VEGFR2 leads to changes in the tumor vasculature that have been termed *vessel normalization*. During the first week of treatment, abnormal vessels are eliminated or pruned (dotted lines), leaving a more normal branching pattern. ECs partially regain features such as cell-cell junctions, adherence to a more normal BM, and pericyte coverage. These changes lead to a decrease in vascular permeability, reduced interstitial pressure, and a transient increase in blood flow within the tumor. Note that in murine models, this normalization period lasts only for ~5–6 days. **D.** After continued anti-VEGF/VEGFR therapy (which is often combined with chemo- or radiotherapy), ECs die, leading to tumor cell death (either due to direct effects of the chemotherapy or lack of blood flow).

pressure within the tumor (which also interferes with the delivery of therapeutics to the tumor; **Figs. 72-9, 72-10, and 72-11**). Tumor blood vessels have a deficit of perivascular cells such as pericytes and smooth-muscle cells that normally regulate flow in response to tissue metabolic needs.

Unlike normal blood vessels, the vascular lining of tumor vessels is not a homogeneous layer of ECs but often consists of a mosaic of ECs and tumor cells, which, because of their plasticity, can upregulate expression of genes normally only seen in ECs under hypoxic conditions. These cancer cell-derived vascular channels, which may be

lined by ECM secreted by the tumor cells, are referred to as *vascular mimicry*. During tumor angiogenesis, ECs are highly proliferative and express a number of plasma membrane proteins that are characteristic of activated endothelium, including growth factor receptors and adhesion molecules such as integrins. These abnormalities in tumor vasculature provide potential differential sensitivities from normal vessels to approaches to inhibit the process, allowing for the use of antiangiogenic agents in cancer treatment.

Lymphatic vessels also exist within tumors. Development of tumor lymphatics is associated with expression of VEGFR3 and its ligands VEGF-C and VEGF-D. The role of these vessels in tumor cell metastasis to regional lymph nodes remains to be determined. However, VEGF-C levels correlate significantly with metastasis to regional lymph nodes in lung, prostate, and colorectal cancers.

■ ANTIANGIOGENIC THERAPY

Angiogenesis inhibitors function by targeting the critical molecular pathways involved in EC proliferation, migration, and/or survival, many of which are highly expressed in the activated endothelium in tumors. Inhibition of growth factor and adhesion-dependent signaling pathways can induce EC apoptosis with concomitant inhibition of tumor growth. Different types of tumors can use distinct combinations of molecular mechanisms to activate the angiogenic switch. Therefore, it is doubtful that a single antiangiogenic strategy will suffice for all human cancers; rather, a number of agents or combinations of agents will be needed, depending on distinct programs of angiogenesis used by different human cancers. Despite this, experimental data indicate that for some tumor types, blockade of a single growth factor (e.g., VEGF) may inhibit tumor-induced vascular growth.

Bevacizumab, an antibody that binds VEGF, potentiates the effects of a number of different types of active chemotherapeutic regimens

used to treat a variety of different tumor types including colon, lung, ovarian, and cervical cancers. It also has activity in combination with interferon against RCCs and alone for glioblastomas. Other protein inhibitors of the VEGF signaling pathway approved for anticancer therapy include ramucirumab (a monoclonal antibody directed against VEGFR2, approved for use against gastric/gastroesophageal, colon, and lung cancers) and ziv-aflibercept (a recombinant protein inhibitor of VEGF, approved for colorectal cancer). Hypertension is the most common side effect of inhibitors of VEGF (or its receptors) but can be treated with antihypertensive agents and uncommonly requires discontinuation of therapy. Rare but serious potential risks include arterial thromboembolic events, including stroke and myocardial infarction, hemorrhage, bowel perforation, and inhibition of wound healing.

Several small-molecule inhibitors (SMIs) that target VEGF RTK activity but are also inhibitory to other kinases have also been approved to treat certain cancers. Sunitinib (see above and Table 72-2) has activity directed against mutant c-Kit receptors (approved for GIST), but also targets VEGFR and PDGFR, and has antitumor activity against pancreatic neuroendocrine and metastatic RCCs, presumably on the basis of its antiangiogenic activity. Similarly, sorafenib, originally developed as a Raf kinase inhibitor but with potent activity against VEGFR and PDGFR, has activity against RCC, differentiated thyroid and hepatocellular cancers, and desmoid tumors. A closely related molecule to sorafenib, regorafenib, has activity against colorectal cancer, GIST, and hepatocellular cancer. Other inhibitors of the VEGF pathway approved for the treatment of various cancers include axitinib, pazopanib, lenvatinib, and cabozantinib.

The success in targeting tumor angiogenesis has led to enhanced enthusiasm for the development of drugs that target other aspects of the angiogenic process; some of these therapeutic approaches are outlined in Fig. 72-12. Recently, an inhibitor of HIF2- α has shown

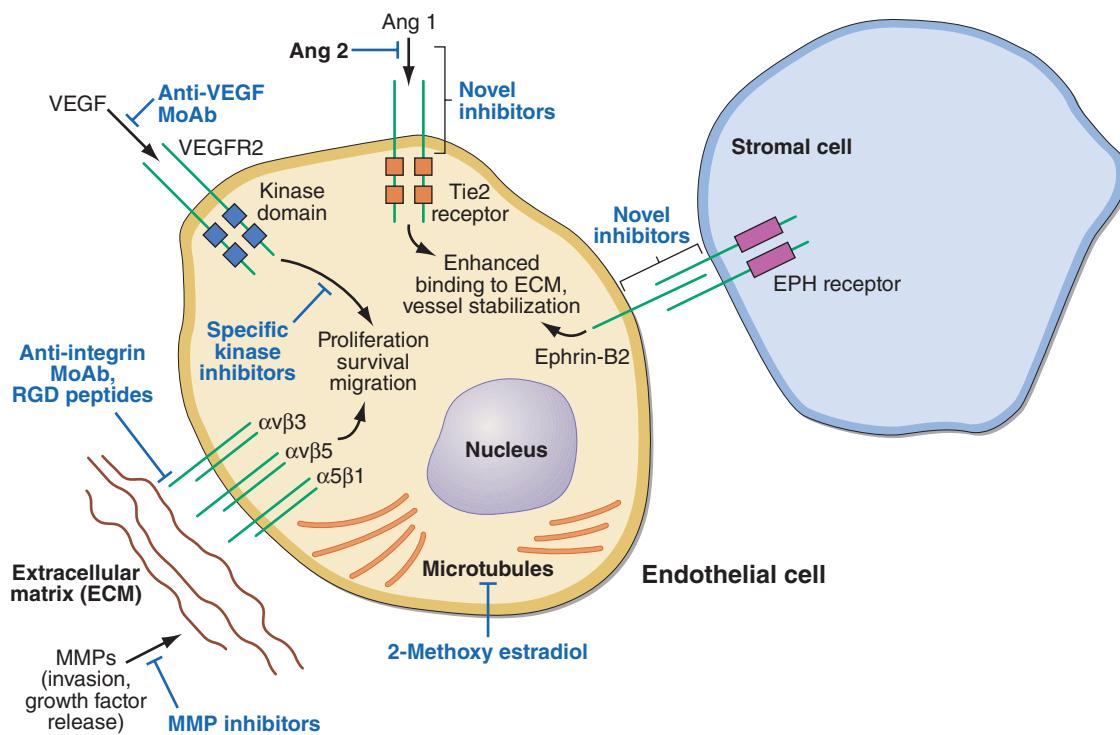


FIGURE 72-12 Knowledge of the molecular events governing tumor angiogenesis has led to a number of therapeutic strategies to block tumor blood vessel formation. The successful therapeutic targeting of VEGF and its receptors VEGFR is described in the text. Other endothelial cell-specific receptor tyrosine kinase pathways (e.g., angiopoietin/Tie2 and ephrin/EPH) are likely targets for the future. Ligation of the $\alpha_v\beta_3$ integrin is required for EC survival. Integrins are also required for EC migration and are important regulators of matrix metalloproteinase (MMP) activity, which modulates EC movement through the ECM as well as release of bound growth factors. Targeting of integrins includes development of blocking antibodies, small peptide inhibitors of integrin signaling, and arg-gly-asp-containing peptides that prevent integrin:ECM binding. Peptides derived from normal proteins by proteolytic cleavage, including endostatin and tumstatin, inhibit angiogenesis by mechanisms that include interfering with integrin function. Signal transduction pathways that are dysregulated in tumor cells indirectly regulate EC function. Inhibition of EGF-family receptors, whose signaling activity is upregulated in a number of human cancers (e.g., breast, colon, and lung cancers), results in downregulation of VEGF and IL-8, while increasing expression of the antiangiogenic protein thrombospondin-1. The Ras/MAPK, PI3K/Akt, and Src kinase pathways constitute important antitumor targets that also regulate the proliferation and survival of tumor-derived EC. The discovery that ECs from normal tissues express tissue-specific "vascular addressins" on their cell surface suggests that targeting specific EC subsets may be possible.

preliminary evidence of antitumor activity against RCC in a clinical trial. There is also evidence suggesting potential enhanced activity when anti-VEGF agents are used in combination with immunomodulators including immune checkpoint inhibitors. However, it is not yet known whether this will produce a clinically meaningful enhancement of antitumor activity.

EVASION OF THE IMMUNE SYSTEM BY CANCERS

The immune system plays a critical role in maintaining organismal integrity including by defending against pathogens as well as preventing and limiting the growth of cancers. There is a complex interaction between cancer and the host from the development of the first malignant cell to the establishment of a clinical cancer and its subsequent growth, invasion, and metastasis. The immune system plays a critical role in the prevention of cancer development. This is exemplified by the increased risk for cancer development in individuals who are significantly immunosuppressed, such as by inherited defects in mechanisms important for immune function, the immunosuppression necessary to maintain allogeneic organ transplants, and immunosuppression seen from certain infections such as human immunodeficiency virus. There are two components of the immune system. The first is innate immunity (present in the organism and not dependent on prior exposure to a specific antigen, such as those present in a pathogen or malignant cell), which tends to be general and not specific. The second is the adaptive immune component, which depends on the innate immune component for activation and provides the specificity to the response with significant expansion of cells to target the specific antigens present on the pathogen or malignant cell. Thus, while the innate process provides the first line of defense, the adaptive process is necessary for the specificity of response and providing memory to more rapidly attack cells should the pathogen infection recur or the malignant cells grow. The immune system has to be tightly regulated to allow for clearance of unwanted antigens while preventing an immune-mediated attack on the self. (See Chap. 349 for details on the function of the immune system).

Not surprisingly, since cancers arise from normal cells within the body that have a variety of processes to prevent destruction by the immune system, they have a variety of mechanisms that allow them to evade detection and elimination by the immune system (Fig. 72-13). These include downregulation of cell surface proteins involved in

immune recognition (including MHC proteins and tumor-specific antigens), expression of other cell surface proteins that inhibit immune function (including members of the B7 family of proteins such as PD-L1), secretion of proteins and other molecules that are immunosuppressive, recruitment and expansion of immunosuppressive cells such as regulatory T cells (which are important for maintaining tolerance against self-antigens), induction of T cell tolerance, and downregulation of death receptors (Fig. 72-14). Due to the marked heterogeneity of cells within a cancer, a variety of immune-suppressive mechanisms are continuously occurring and changing. In addition, the inflammatory effects of some of the immune mediator cells in the tumor microenvironment (especially tissue-associated macrophages and myeloid-derived suppressor cells) can suppress T cell responses to the tumor as well as stimulate inflammation that can enhance tumor growth.

There are marked differences in the way different malignancies respond to current immunotherapeutic approaches. For example, melanomas, RCC, Merkel cell carcinomas, cancers with defects in DNA repair associated with microsatellite instability with accumulation of gene mutations, and lymphomas respond well to current immunotherapeutic approaches, whereas pancreatic and microsatellite-stable colon cancers do not. While there is not a complete understanding of why these differences exist and there are many factors both within the cancer cells and in the microenvironment that play a role, several factors have been identified that appear to be important. These include the number of mutations present in the tumor (tumor mutational burden), presence of new or neoantigens, expression of immune checkpoint proteins (e.g., PD-L1 for anti-PD-1 or anti-PD-L1 therapy), density of tumor-infiltrating lymphocytes, and host genetic factors. One of these (PD-L1 expression by the tumor) has sufficient predictive value for certain tumors (e.g., non-small-cell lung cancer) to be used in making treatment decisions regarding the use of antibodies targeting PD-1 or PD-L1. However, neither PD-L1 expression nor any other marker can predict responsiveness of most tumors to immunotherapy. Better biomarkers that define potential responsiveness of specific cancers to immunotherapy are badly needed. A major area of research is to try to identify approaches that would convert cancers that are not responsive to immunotherapy to being responsive.

Immunotherapy approaches to treat cancer can be divided into those aimed at activating the immune response and those designed to

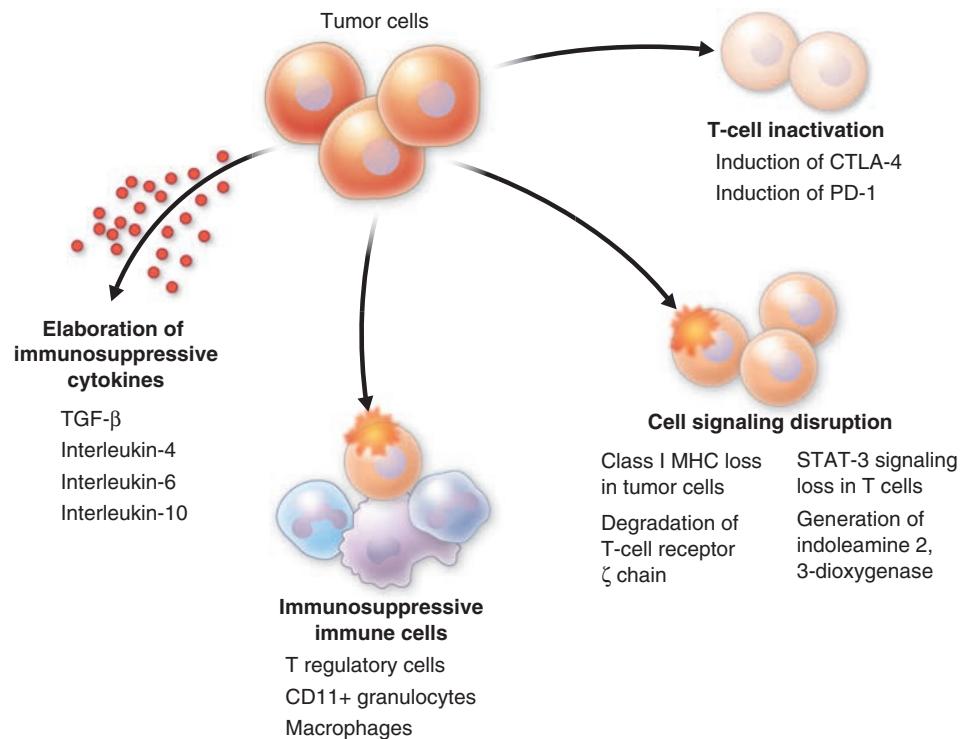


FIGURE 72-13 Tumor-host interactions that suppress the immune response to the tumor.

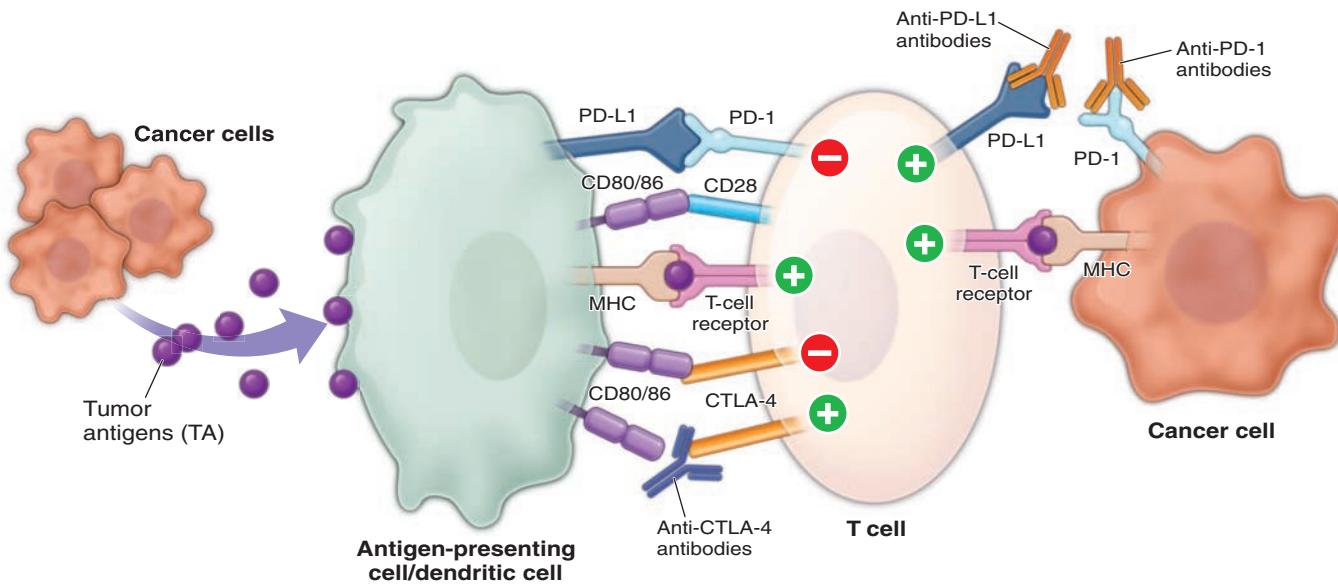


FIGURE 72-14 Inhibition of T-cell activation against cancer cells by engagement of co-inhibitory molecules including PD-1, PD-L1, and CTLA-4 and reversal of this inhibition by antibodies against these proteins. The red ovals in the T cell indicate inhibitory signals, and the green oval indicates stimulatory signals.

release the brakes that prevent an effective immune response against tumors. Approaches at activating the immune response against cancer including using immunostimulatory molecules such as interferons, IL-2, and especially monoclonal antibodies have had some success.

A more direct approach to enhance the activity of T cells directed against specific tumors involves isolating T cells from patients and re-engineering the cells to express chimeric antigen receptors (CAR-T cells) that recognize antigens present on the cells of that individual's tumor. The most commonly used approach to date has been to engineer the cells to express receptors targeting the CD19 antigen on acute lymphocytic leukemia (ALL) and diffuse large B-cell lymphoma (DLBCL) cells. These have been shown to have significant antitumor activity in the treatment of patients with ALL and DLBCL, including durable remissions in patients refractory to standard therapies, and are approved for these malignancies. However, there have also been significant issues with toxicity including cytokine release syndrome, organ toxicity felt to be due to inadvertent targeting of antigens present in the organ, and neurotoxicity. These patients often require aggressive supportive care by individuals experienced in the delivery of CAR-T cells. In addition, as is true for most anticancer therapies, mechanisms of resistance have developed, most commonly the outgrowth of tumor cells no longer expressing the antigen. Mechanisms for preventing the development of resistant cells are being explored. CAR-T cells are undergoing clinical investigation against other hematologic malignancies (e.g., multiple myeloma) and solid tumors. Approaches to develop allogeneic CAR-T-cell therapies are also being explored.

The other approach to enhancing the immune response against cancers is releasing the brakes that inhibit a response by targeting of proteins or cells (e.g., regulatory T cells) involved in normal homeostatic control to prevent autoimmune damage to the host but that malignant cells and their stroma can also utilize to inhibit the immune response directed against them. The approach that is furthest along clinically has involved targeting CTLA-4, PD-1, and PD-L1 (and others)—co-inhibitory molecules that are expressed on the surface of cancer cells, cells of the immune system, and/or stromal cells and are involved in inhibiting the immune response against cancer (Figs. 72-13 and 72-14). This approach has had clinical activity against a variety of cancers. A monoclonal antibody directed against CTLA-4 is approved for the treatment of melanoma, and antibodies targeting PD-1 or PD-L1 are approved for use against many cancers, including melanoma, RCC, lung cancer (both non-small-cell lung and small-cell lung), head and neck cancer, urothelial cancer, cervical cancer, hepatocellular carcinoma, gastric cancer, esophageal cancer, microsatellite instability (MSI)-high cancers, cancers with high tumor mutational burden

(TMB), Merkel cell cancer, primary B-cell mediastinal lymphoma, and Hodgkin's lymphoma. They continue to be evaluated against other malignancies as well. The combination of anti-CTLA-4 and anti-PD-1 antibodies has been approved for treatment of a number of cancers including melanoma, RCC, lung cancer, pleural mesothelioma, and MSI-high metastatic colorectal cancers. Immune checkpoint inhibitors are being used singly, in pairs, and in combination with chemotherapy in many ongoing clinical trials. Specific determinants of response to immune checkpoint inhibitors are still being defined, but in addition to high PD-L1 expression, the presence of increased neoantigens in the tumor, such as seen in patients with MSI-high and TMB-high cancers, may be one important determinant of better responses.

A number of other proteins are involved in controlling the immune response (both ones that enhance activity [e.g., CD27 and CD40] as well as ones involved in inhibiting response [e.g., LAG3, TIM-3, TIGIT]). Antibodies have been developed to modulate function of these proteins, and many are in clinical development for cancer therapy. In addition, various combinations targeting more than one protein involved in potentially enhancing the immune response against cancers or with other anticancer approaches (targeted agents, chemotherapy, radiation therapy) that may lead to enhanced antitumor activity are also being explored. An important aspect of these approaches is balancing sufficient release of the negative control of the immune response to allow immune-mediated attack on the tumors while not allowing too much of an immune response against normal tissues and thus inducing severe autoimmune effects (e.g., against lung, liver, skin, thyroid, pituitary gland, or the gastrointestinal tract). As is true for other immunotherapeutic approaches against cancer, major efforts are ongoing to better understand the mechanism of immune toxicity from these approaches and, therefore, ways of controlling this while not abrogating the antitumor effects.

Improved knowledge of the biology of the interactions between the immune system and cancers continues to be rapid with the promise for additional significant improvements in use of immunotherapy to treat cancer.

SUMMARY

Although each of the biological aspects of cancers and examples of targeting them has been addressed individually, clearly there is complicated cross-talk between these that occurs in all cancers that needs to be better understood to optimally treat different cancers. The explosion of information on tumor cell biology, metastasis, and tumor-host interactions (including angiogenesis, other tumor-stromal interactions, and immune evasion by tumors) has ushered in a new era of rational

targeted therapy for cancer. Furthermore, it has become clear that specific molecular factors detected in individual tumors (specific gene mutations, gene expression profiles, miRNA expression, overexpression of specific proteins) can be used to tailor therapy and maximize antitumor effects. Potentially of greater impact on decreasing deaths from cancer, better understanding of the biology of early cancer development and technologic development to improve sensitivity and specificity in detecting cancer-specific molecules (e.g., mutated genes) provide hope that approaches for earlier detection of cancer can be developed.

ACKNOWLEDGMENT

Robert G. Fenton contributed to this chapter in prior editions, and important material from those prior chapters has been included here.

■ FURTHER READING

- Boussioutis VA: Molecular and biochemical aspects of the PD-1 checkpoint pathway. *N Engl J Med* 375:1767, 2016.
- De Palma M et al: Microenvironmental regulation of tumour angiogenesis. *Nat Rev Cancer* 17:457, 2017.
- Du W, Elemento O: Cancer systems biology: Embracing complexity to develop better anticancer therapeutic strategies. *Oncogene* 34:3215, 2015.
- Fleuren ED et al: The kinase “at large” in cancer. *Nat Rev Cancer* 16:83, 2016.
- He S, Sharpless NE: Senescence in health and disease. *Cell* 169:1000, 2017.
- Lambert AW et al: Emerging biological principles of metastasis. *Cell* 168:670, 2017.
- Otto T, Sicinski P: Cell cycle proteins as promising targets in cancer therapy. *Nat Rev Cancer* 17:93, 2017.
- Tomasetti C et al: Stem cell divisions, somatic mutations, cancer etiology and cancer prevention. *Science* 355:1330, 2017.
- Vander Heiden MG, DeBerardinis RJ: Understanding the intersections between metabolism and cancer biology. *Cell* 168:657, 2017.
- Vogelstein B et al: Cancer genome landscapes. *Science* 339:1546, 2013.

TABLE 73-1 Spectrum of Cancer-Related Interventions

Asymptomatic patient (breast, cervix, colon, some lung) screening
Consideration of cancer in a differential diagnosis
Physical examination, imaging, or endoscopy to define a possible tumor
Phlebotomy for molecular studies and circulating tumor cell characterization
Diagnosis of cancer by biopsy or removal:
Routine histology
Specialized histology: immunohistochemistry
Molecular studies
Cytogenetic studies
Staging the cancer: Where has it spread?
Treatment
Localized (surgical removal with or without local radiation therapy and/or topical therapy may be curative)
Systemic (prevent or reverse organ compromise)
Supportive care
During treatment: related to tumor effects on patient
During treatment: to counteract side effects of treatment
After treatment: to ameliorate the adverse effects of treatment
Palliative and end of life
When useful treatments are not feasible or desired

of hollow viscera, but also may reflect altered platelet number or blood coagulation. Tumors may also present with a “paraneoplastic syndrome” owing to the effects of substances they secrete. Although statistically the fraction of patients with cancer underlying a particular presenting sign or symptom may be low, the implications of missing an early-stage tumor call for vigilance in considering cancer as the basis for persistent signs or symptoms.

Evidence of a tumor’s existence can come from careful physical examination, e.g., enlarged lymph nodes in lymphomas or palpable mass in a breast or soft tissue site. A mass may also be detected or confirmed by an imaging modality, such as plain x-ray, computed tomography (CT) scan, ultrasound, positron emission tomography (PET) imaging, or nuclear magnetic resonance approaches. Endoscopy may directly visualize a tumor.

■ ESTABLISHING A CANCER DIAGNOSIS

Once a potential tumor is defined, establishing the diagnosis is the next step in the intervention spectrum. This requires a biopsy procedure in most circumstances and pathologic confirmation that cancer is present; very rarely, where biopsy would be definitely injurious and imaging modalities are unequivocal, such as with a likely brainstem glioma, treatment might be reasonably considered based on clinical and imaging evidence without biopsy. In addition to light microscopy, biopsied tissue also allows definition of genetic abnormalities and protein expression patterns (**Table 73-2**).

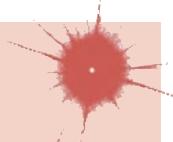
The extent of specialized testing needs to be tailored to an individual patient’s case. Global DNA sequencing of genes expressed in tumors has not been shown to convey conclusive advantage in terms of survival. But the aggregate “mutational burden” present in tumors and the intactness of DNA repair genes (e.g., breast cancer susceptibility 1 and 2 [*BRCA1/2*], microsatellite instability, homologous recombination pathway-associated genes) may suggest valuable treatment courses in tumors without curative potential. Testing for certain abnormalities in Table 73-2 can be the basis for use of specific U.S. Food and Drug Administration (FDA)-approved therapeutic agents.

Optimally, an *excisional biopsy* occurs, in which the entire tumor mass is removed with a margin of normal tissue surrounding it. If an excisional biopsy cannot be performed, *incisional biopsy* is the procedure of second choice: a wedge of tissue is removed, trying to include the majority of the cross-sectional diameter to minimize sampling error. Biopsy techniques that involve cutting into tumor risk facilitating the spread of the tumor, and consideration with a surgeon of whether the biopsy approach is a potential prelude to a curative surgery

73

Principles of Cancer Treatment

Edward A. Sausville, Dan L. Longo



CANCER PRESENTATION

Localized or systemic cancer is frequent in the differential diagnosis of a variety of common complaints. Affording patients the greatest opportunity for cure or meaningful prolongation of life is greatly aided by cancer diagnosis early in its natural history. The spectrum of possible cancer-related interventions to make cure possible are shown in **Table 73-1**.

■ DETECTION OF A CANCER

The term *cancer*, as used here, is synonymous with the term *tumor*, whose original derivation from Latin simply meant “swelling,” not otherwise specified. Swelling reflects increased interstitial fluid pressure and increased cellular and stromal mass, compared to normal tissue. Leukemias, a cancer of the blood-forming tissues, presents in a disseminated form frequently without tumor masses. Tumors can also present by organ dysfunction, such as dyspnea on exertion from anemia caused by leukemia replacing normal marrow, cough from lung cancers, jaundice from tumors blocking bile ducts, or neurologic signs from gliomas. Hemorrhage frequently results from involvement

TABLE 73-2 Diagnostic Biopsy: Standard-of-Care Molecular and Special Studies to Be Considered

All solid tumors:
Tumor mutational burden
Microsatellite instability DNA repair pathway intactness
Homologous recombination DNA repair pathway intactness
Breast cancer: primary and suspected metastatic
Breast cancer susceptibility 1 and 2 (<i>BRCA1/2</i>) gene mutations
Hormone receptor expression: estrogen, progesterone
HER2/neu oncoprotein
<i>PI3KA</i> mutation status
Lung cancer: primary and suspected metastatic
If nonsquamous non-small-cell:
Epidermal growth factor receptor (<i>EGFR</i>) mutation
<i>ALK</i> gene fusion
<i>BRAFV600E</i> mutation
Programmed cell death ligand 1 (PD-L1) expression
Colon cancer: suspected metastatic
<i>KRAS</i> mutation
<i>BRAFV600E</i> mutation
Gastrointestinal stromal tumor
<i>KIT</i> mutation
Melanoma
<i>BRAF</i> mutation
c-kit expression and <i>KIT</i> mutation if present
Pancreatic cancer
<i>BRCA1/2</i> mutation
Prostate cancer
<i>BRCA1/2</i> mutation
Thyroid cancer
<i>RET</i> gene alterations (mutations, translocations, amplification)
Gliomas
1p/19q co-deletion
Alkylguanine alkyltransferase promoter methylation
Isocitrate dehydrogenase 1 and 2 mutation
Leukemia (peripheral blood mononuclear cells and/or bone marrow)
Cytogenetics
Flow cytometry
Treatment-defining chromosomal translocations/mutations
Bcr-Abl fusion protein
t(15;17)
inversion 16
t(8;21)
FMS-associated tyrosine kinase (<i>FLT3</i>) mutation
Nucleophosmin gene mutational status
Isocitrate dehydrogenase 1 and 2 mutation
Lymphoma
Immunohistochemistry for CD20, CD30, T-cell markers
Treatment-defining chromosomal translocations:
t(14;18)
t(8;14)
Translocations involving <i>ALK</i> gene

accounting for possible diagnoses may best inform the approach taken. *Core-needle biopsy* usually obtains considerably less tissue but can provide information to plan a treatment. *Fine-needle aspiration* generally yields a suspension of cells from a mass. If positive for cancer, it may allow inception of systemic treatment, or it can provide a basis for planning a more extensive surgical procedure. It is unreliable as a sole diagnostic method to make a cancer diagnosis in most cases. A “negative” fine-needle aspiration cannot be taken as definitive evidence that a tumor is absent. In some instances, features of diagnostic imaging are sufficient to make a reliable diagnosis without obtaining tissue, usually

with presence of a tumor associated circulating diagnostic marker, e.g., alpha fetoprotein in hepatocellular carcinoma.

CANCER STAGING

An essential component of correct patient management in many cancer types is defining the extent of disease to determine whether localized treatments, “combined-modality” approaches, or systemic treatments should initially be considered. Radiographic and other imaging tests can be helpful in defining the *clinical stage*; *pathologic staging* documents the histologic presence of tumor in tissue biopsies obtained through a surgical procedure. Lymph node sampling in breast cancer, melanoma, lung, head and neck, colon, and other intra-abdominal cancers may provide crucial information.

Staging systems have evolved to define a “T” component related to the size of the tumor or its invasion into local structures, an “N” component related to the number and nature of lymph node groups adjacent to the tumor with evidence of tumor spread, and an “M” component, based on the presence of local or distant metastatic sites. The various TNM components are then aggregated to stages, usually stage I to III or IV, depending on the anatomic site. The numerical stages reflect similar long-term survival outcomes of the aggregated TNM groupings in a numeric stage after treatment tailored to the stage. In general, stage I tumors are T1 (reflecting small size), N0 or N1 (reflecting no or minimal node spread), and M0 (no metastases). Such early-stage tumors are usually amenable to curative approaches with local treatments. On the other hand, stage IV tumors have metastasized to distant sites or locally invaded viscera in a nonresectable way. They are treated with palliative intent, except for those diseases with exceptional sensitivity to systemic treatments such as chemotherapy or immunotherapy. Also, the TNM staging system is not useful in diseases such as leukemia, where bone marrow infiltration is never localized, or central nervous system (CNS) tumors, where tumor histology and the extent of feasible resection are more important in driving prognosis.

CANCER TREATMENT

The goal of cancer treatment is first to eradicate the cancer; if not possible, the goal shifts to palliation: amelioration of symptoms and preservation of quality of life while striving to extend life. When cure of cancer is possible, cancer treatments may be undertaken despite the certainty of severe toxicities, and these may produce toxicity with no benefit. Conversely, when the clinical goal is palliation, careful attention to minimizing the toxicity of treatments becomes a significant goal.

Cancer treatments are divided into two main types: *local* and *systemic*. Local treatments include surgery, radiation therapy (including photodynamic therapy), and ablative approaches, including radiofrequency and thermal or cryosurgical approaches. Systemic treatments include chemotherapy (including hormonal therapy and molecular targeted therapy) and biologic therapy (including immunotherapy). The modalities are often used in combination. *Oncology*, the study of tumors including treatment approaches, is a multidisciplinary effort with surgical, radiation, and internal medicine-related areas of oncologic expertise.

Normal organs and cancers share the property of having a population of cells actively progressing through the cell cycle, with their division providing a basis for organ or tumor growth, and a population of cells not in cycle; these include *stem cells*, whose properties are being elucidated. Cancer stem cells serve as a basis for tumor initiating or repopulating cells. Tumors follow a Gompertzian growth curve (**Fig. 73-1**), with the growth fraction of a neoplasm high with small tumor burdens and declining until, at the time of diagnosis, with a tumor burden of $1\text{--}5 \times 10^9$ tumor cells, the growth fraction is usually 1–4% for many solid tumors. By this view, the most rapid growth rate occurs before the tumor is detectable. An alternative explanation for such growth properties may also emerge from the ability of tumors at metastatic sites to recruit circulating tumor cells from the primary tumor or other metastases. Key features of tumor growth are the ability to stimulate new supporting stroma through angiogenesis and ingrowth of fibroblasts and immune cells (**Chap. 72**).

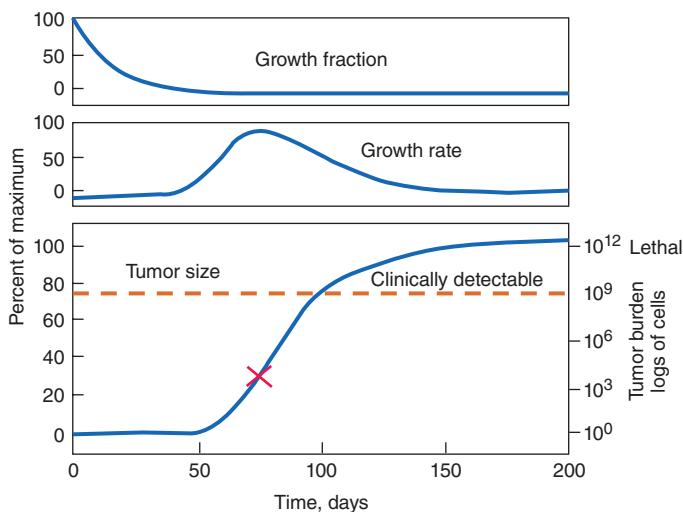


FIGURE 73-1 Gompertzian tumor growth. The growth fraction of a tumor declines exponentially over time (top), peaking before it is clinically detectable (middle). Tumor size increases slowly, goes through an exponential phase, and slows again as the tumor has limitation of nutrients or host regulatory influences occur. The maximum growth rate occurs at $1/e$, the point at which the tumor is about 37% of its maximum size (marked with an X). Tumor becomes detectable at a burden of about 10^9 (1 cm^3) cells and kills the patient at a tumor cell burden of about 10^{12} (1 kg).

LOCALIZED CANCER TREATMENTS

SURGERY

Surgery is unquestionably the most effective means of treating cancer. At least 40% of cancer patients are cured by surgery. Unfortunately, a large fraction of patients with solid tumors have metastatic disease not accessible for removal. Even when cancer is not curable by surgery alone, the removal of tumor can afford local control of tumor, preserve organ function, achieve debulking that permits more effective subsequent therapy, and allow more detailed staging. Cancer surgery aiming for cure is usually planned to excise the tumor completely with an adequate margin of normal tissue (the margin varies with the tumor and the anatomy), touching the tumor as little as possible to prevent vascular and lymphatic spread, and minimizing operative risk. Such a resection is defined as an R0 resection. R1 and R2 resections, in contrast, are imprecisely defined pathologically as having microscopic or macroscopic, respectively, tumor at resection margins. Such outcomes may be the basis for reoperation to obtain optimal margins if feasible and of likely clinical utility. Extending the procedure to resect draining lymph nodes obtains prognostic information and may, in some anatomic locations, improve survival.

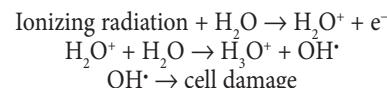
Laparoscopic approaches are being used for primary abdominal and pelvic tumors, although with certain tumors (e.g., uterine and cervix), controversy exists as to the desirability of laparoscopic tissue removal. Lymph node spread may be assessed using the sentinel node approach, in which the first draining lymph node is defined by injecting a dye or radioisotope into the tumor site at operation and then resecting the first node to turn blue or collect isotope. The sentinel node assessment appears to provide information without the risks (lymphedema, lymphangiosarcoma) associated with resection of all regional nodes. Advances in adjuvant chemotherapy (chemotherapy given systemically after removal of all local disease surgically without evidence of active metastatic disease) and radiation therapy following surgery have permitted a substantial decrease in the extent of primary surgery necessary to obtain the best outcomes. Thus, “lumpectomy” with radiation therapy is as effective as modified radical mastectomy for breast cancer, and limb-sparing surgery followed or preceded by adjuvant radiation therapy and chemotherapy has replaced amputation for most childhood rhabdomyosarcomas and osteosarcomas. More limited surgery spares organ function, as in larynx and bladder cancer. In some settings (e.g., bulky testicular cancer or stage III breast cancer), surgery is not the first treatment modality used. After diagnostic biopsy, chemotherapy and/or radiation therapy is delivered, followed by a surgical procedure

to remove residual masses; this is called *neoadjuvant therapy*. Coordination among the surgical oncologist, radiation oncologist, and medical oncologist is crucial.

Surgery may be curative in a subset of patients with metastatic disease. Patients with limited lung metastases from osteosarcoma may be cured by resection of the lung lesions. In patients with colon cancer who have fewer than five liver metastases restricted to one lobe and no extrahepatic metastases, hepatic lobectomy may produce long-term disease-free survival in 25% of selected patients. In the setting of hormonally responsive tumors, oophorectomy may eliminate estrogen production, and orchietomy may reduce androgen production, hormones that drive metastatic breast and all prostate cancers, respectively. In selecting a surgeon or center for primary cancer treatment, consideration must be given to the volume of cancer surgeries undertaken by the site. Studies in a variety of cancers have shown that increased annual procedure volume appears to correlate with outcome. Surgery is used in a number of ways for palliative or supportive care of the cancer patient. These include insertion and care of central venous catheters, control of pleural and pericardial effusions and ascites, caval interruption for recurrent pulmonary emboli, stabilization of cancer-weakened weight-bearing bones, and control of hemorrhage, among others. Surgical bypass of gastrointestinal, urinary tract, or biliary tree obstruction can alleviate symptoms and prolong survival. Surgical procedures may provide relief of pain or neurologic dysfunction (spinal cord decompression). Splenectomy may relieve symptoms and reverse hypersplenism. Intrathecal or intrahepatic therapy relies on surgical placement of appropriate infusion portals. Surgery may correct other treatment-related toxicities such as adhesions or strictures. Plastic and reconstructive surgery can correct the effects of disfiguring primary treatment. Surgery is also a tool valuable in the prevention of cancers in high-risk populations. Prophylactic mastectomy, colectomy, oophorectomy, and thyroidectomy are mainstays of prevention of genetic cancer syndromes.

RADIATION

Radiation Biology and Medicine Therapeutic radiation is ionizing, causing breaks in DNA and generation of free radicals from cell water that damage cancer cell membranes, proteins, and organelles. Radiation damage is augmented by oxygen; hypoxic cells are more resistant.



X-ray and gamma-ray photons are the forms of ionizing radiation most commonly used to treat cancer. Particulate ionizing radiation using protons has also become available.

Radiation dose is quantitated based on the amount of energy absorbed by the tumor, not on radiation generated by the machine. The International System (SI) unit for radiation dose is the Gray (Gy): 1 Gy refers to 1 J/kg of tissue; 1 Gy equals 100 centigrays (cGy) of absorbed dose. A historically used unit appearing in the oncology literature, the *rad* (radiation absorbed dose), is defined as 100 ergs of energy absorbed per gram of tissue and is equivalent to 1 cGy. Radiation dose is measured by placing detectors at the body surface or in irradiated phantoms that resemble human form and substance. The features that make a particular cell more or less sensitive to radiation involve DNA repair proteins that, in their physiologic role, protect against environmentally related DNA damage.

Localized Radiation Therapy Radiation effect is influenced by three determinants: total absorbed dose, number of fractions, and time of treatment. A typical course of radiation therapy should be described as 4500 cGy delivered to a particular target (e.g., mediastinum) over 5 weeks in 180-cGy fractions. Most curative radiation treatment programs are delivered once a day, 5 days a week, in 150- to 200-cGy fractions. Nondividing cells are more resistant than dividing cells; delivering radiation in repeated fractions is done to expose a larger

number of tumor cells that have entered the division cycle. The energy of the radiation determines its ability to penetrate tissue. Low-energy x-rays (150–400 kV) scatter when they strike the body, resulting in more damage to adjacent normal tissues and less radiation delivered to the tumor. Megavoltage radiation (>1 MeV) has very low lateral scatter; this produces a skin-sparing effect, more homogeneous distribution of the radiation energy, and greater deposit of the energy in the tumor, or *target volume*. The *transit volume* includes the tissues through which the beam passes to the target volume. Computational approaches and delivery of many beams to converge on a target volume are the basis for “gamma knife” and related approaches to deliver high doses to tumor, sparing normal tissue.

Therapeutic radiation is delivered in three ways: (1) *teletherapy*, with focused beams of radiation generated at a distance and aimed at the tumor within the patient; (2) *brachytherapy*, with encapsulated sources of radiation implanted directly into or adjacent to tumor tissues; and (3) *systemic therapy*, with radionuclides administered, for example, intravenously but perhaps targeted by some means to a tumor site. Teletherapy with x-ray or gamma-ray photons is the most commonly used form of radiation therapy and also delivers particulate forms of radiation such as proton beams. The difference between photons and protons relates to volume with greatest delivery of energy: protons have a narrow range of energy deposition. Electron beams are a particulate form of radiation that, in contrast to photons and protons, have a very low tissue penetrance and are used to treat cutaneous tumors. Certain drugs used in cancer treatment may also act as radiation sensitizers. For example, compounds that incorporate into DNA (e.g., halogenated pyrimidines, cisplatin) augment radiation effects at local sites and are important adjuncts to radiation of certain tumors, e.g., squamous head and neck, uterine cervix, and rectal cancers.

Toxicity of Radiation Therapy Although radiation therapy is most often administered to a local region, systemic effects, including fatigue, anorexia, nausea, and vomiting, may develop that are related in part to the volume of tissue irradiated, dose fractionation, radiation fields, and individual susceptibility. Injured tissues release cytokines that act systemically to produce these effects. Bone is among the most radio-resistant organs, with radiation effects being manifested mainly in children through premature fusion of the epiphyseal growth plate. By contrast, the male testis, female ovary, and bone marrow are the most sensitive organs. Any bone marrow in a radiation field will be eradicated by therapeutic irradiation. Organs with less need for cell renewal, such as heart, skeletal muscle, and nerves, are more resistant to immediate radiation effects. In radiation-resistant organs, the vascular endothelium is the most sensitive component. Acute toxicities include mucositis, skin erythema (ulceration in severe cases), and bone marrow toxicity. Often these can be alleviated by periodic interruption of treatment.

Chronic toxicities are more serious. Radiation of the head and neck region produces thyroid failure; cataracts and retinal damage can lead to blindness; salivary glands stop making saliva, which leads to dental caries and poor dentition. Mediastinal irradiation increases myocardial vascular disease. Other late vascular effects include chronic constrictive pericarditis, lung fibrosis, viscous stricture, spinal cord transection, and radiation cystitis or enteritis.

A serious late toxicity is the development of second solid tumors in or adjacent to the radiation fields. Such tumors can develop in any organ or tissue and occur at a rate of ~1% per year beginning in the second decade after treatment.

■ OTHER LOCALIZED CANCER TREATMENTS

Endoscopy allows placement of stents to unblock viscera by mechanical means, palliating, for example, gastrointestinal or biliary obstructions. Radiofrequency ablation (RFA) refers to focused microwave nonionizing radiation to induce thermal injury within a volume of tissue. RFA can be useful in the control of metastatic lesions, particularly in liver, that may threaten biliary drainage (as one example in patients with otherwise unresectable disease). Cryosurgery uses extreme cold to sterilize lesions in certain sites, such as prostate and kidney, at a very

early stage, eliminating the need for modalities with more side effects such as surgery.

Some chemicals (porphyrins, phthalocyanines) are preferentially taken up by cancer cells. When intense light, delivered by a laser, is shone on cells containing these compounds, free radicals are generated and the cells die. Such phototherapy is used to treat skin cancer; ovarian cancer; and cancers of the lung, colon, rectum, and esophagus. Palliation of recurrent locally advanced disease can sometimes be dramatic and last many months.

Infusion of chemotherapeutic or biologic agents or radiation-bearing delivery devices such as isotope-coated glass spheres into local sites through catheters have been used to treat disease limited to that site; in selected cases, prolonged control of truly localized disease has been possible.

SYSTEMIC CANCER TREATMENTS

The concept that systemically administered chemicals might have a useful effect on cancers was historically derived from three sets of observations. Paul Ehrlich in the nineteenth century observed that different dyes reacted with different cell and tissue components. He hypothesized the existence of “magic bullets” that might bind to tumors, owing to the affinity of the agent for the tumor. Observation of the toxic effects of certain mustard gas derivatives on the bone marrow during World War I suggested that smaller doses of these agents might be used to treat tumors of marrow-derived cells. Finally, the fact that tumors from hormone-responsive tissues, e.g., breast tumors, could shrink after oophorectomy led to the idea that endogenous or exogenous substances might modulate tumor growth by altering its regulatory biology. Chemicals achieving each of these goals are currently used as cancer chemotherapy agents.

Anecdotal reports of tumor regression following intratumoral injection of bacterial extracts raised the possibility of immune system-mediated tumor regression. Serotherapy of infectious disease in the preantibiotic era encouraged analogous efforts to develop vaccine- and antibody-based treatments for cancer. Administration of autologous immune cells obtained by pheresis procedures from a patient or purified from a patient’s removed tumor, activated by cytokines *ex vivo*, achieved durable disease control in a small fraction of patients. These observations provided the rationale for more modern efforts to treat tumors using cell-mediated immunity.

Systemic cancer treatments are of three broad types. *Cytotoxic chemotherapy agents* are “small molecules” (generally with molecular mass <1500 Da) that cause major regression of experimental tumors growing in animals. These agents mainly target DNA structure or segregation of chromosomes in mitosis. *Cancer molecular target therapies* refer to small molecules designed and developed to interact with a defined macromolecule important in maintaining the malignant state. As described in Chap. 72, successful tumors have activated biochemical pathways that lead to uncontrolled proliferation through the action of hormone receptor proteins, oncogene products, loss of cell cycle inhibitors, or loss of cell death regulation, and have acquired the capacity to replicate chromosomes indefinitely, invade, metastasize, and evade the immune system. *Cancer biologic therapies* are most frequently macromolecules, cells, or cell extracts that have a particular target (e.g., anti-growth factor receptor, cytokine, or immunomodulatory antibodies) or may have the capacity to induce a host immune response to kill tumor cells. Most recent additions to cancer biologic therapies include genetically modified cells that directly attack tumor cells and tumor-infecting viruses that can kill tumor cells but also elicit host antitumor immune responses.

■ SYSTEMIC CANCER THERAPY OVERVIEW

General Principles The *therapeutic index* of any drug is the degree of separation between toxic and therapeutic doses. Really useful drugs have large therapeutic indices, and this usually occurs when the drug target is expressed in the disease-causing compartment as opposed to the normal compartment. Cytotoxic chemotherapeutic agents have the unfortunate property that their main targets, DNA and microtubules,

are present in both normal and tumor tissues. Therefore, they have relatively narrow therapeutic indices. Targeted agents can also cause effects on their target in normal tissues, or “off-target” effects on unrelated targets in organs experiencing damage. Biologic therapies may elicit misdirected immune responses on normal organ function. A key activity in oncology drug development is striving to administer a dose of agent that can convey benefit with a minimal or tolerable side effect profile.

Figure 73-2 illustrates steps in cancer drug development. Following demonstration of antitumor activity in animal models, potentially useful anticancer agents are further evaluated to define an optimal schedule of administration and suitable drug formulation. Safety testing in two animal species on an analogous schedule of administration defines the starting dose for a phase 1 trial in humans, usually but not always in patients with cancer who have exhausted “standard” (already approved) treatments. The initial dose is usually one-sixth to one-tenth of the dose just causing easily reversible toxicity in the more sensitive animal species. If the agent is not intrinsically toxic, doses of drug achieving fractions of the useful concentration from model systems are studied. Escalating doses of drug are then given during the human phase 1 trial until reversible toxicity is observed or the desired drug concentration is achieved. Dose-limiting toxicity (DLT) defines a dose that conveys greater toxicity than would be acceptable in routine practice, allowing definition of a lower maximum-tolerated dose (MTD). The occurrence of toxicity is, if possible, correlated with plasma drug concentrations. The MTD or a dose just lower than the MTD is usually the dose suitable for phase 2 trials, where a fixed dose is administered to a relatively homogeneous set of patients with a particular tumor type. If no toxicity has emerged in phase 1 trials, administration of the optimal biologic dose to achieve effective drug concentrations is undertaken. A partial response (PR) historically was defined as a decrease of at least 50% in a tumor’s bidimensional area obtained by imaging; more recent response criteria (e.g., Response Evaluation Criteria in Solid Tumors

[RECIST]) may use a 30% decrease in aggregate unidimensional areas of target lesions. Response criteria for immunologically directed agents may allow a substantial transient increase in tumor volume as long as a patient’s clinical status is stable, as these agents may evoke inflammatory responses in tumors with subsequent shrinkage or stabilization of lesions then occurring subsequently. A complete response (CR) connotes disappearance of all tumor; progression of disease signifies an increase in size of existing lesions by >25% from baseline or best response or development of new lesions; stable disease fits into none of the above categories.

In a phase 3 trial, evidence of improved overall survival or improvement in the time to progression of disease on the part of the new drug is sought in comparison to an appropriate control population. Data from the entire process are the basis for application to a regulatory agency to approve the new agent for commercial marketing.

Cancer drug clinical trials conventionally use a toxicity grading scale where grade 1 toxicities do not require treatment, grade 2 toxicities may require symptomatic treatment but are not life-threatening, grade 3 toxicities are potentially life-threatening if untreated, grade 4 toxicities are actually life-threatening, and grade 5 toxicities are those that result in the patient’s death. Active efforts to quantitate effects of anticancer agents on quality of life also frequently occur in early development of oncology drugs.

Development of targeted agents may proceed differently. While phase 1–3 trials are still conducted, focus on a particular tumor type even in phase 1 may be enabled by molecular analysis to define target expression in a patient’s tumor necessary for or relevant to the drug’s action. Ideally, pharmacodynamic studies would also assess whether the target has been hit. The failure of a targeted therapy can be either because the drug missed the target or it hit the target but the target was not central to the tumor’s growth and survival. Within the past decade, agents have been approved for clinical use not in relation to an

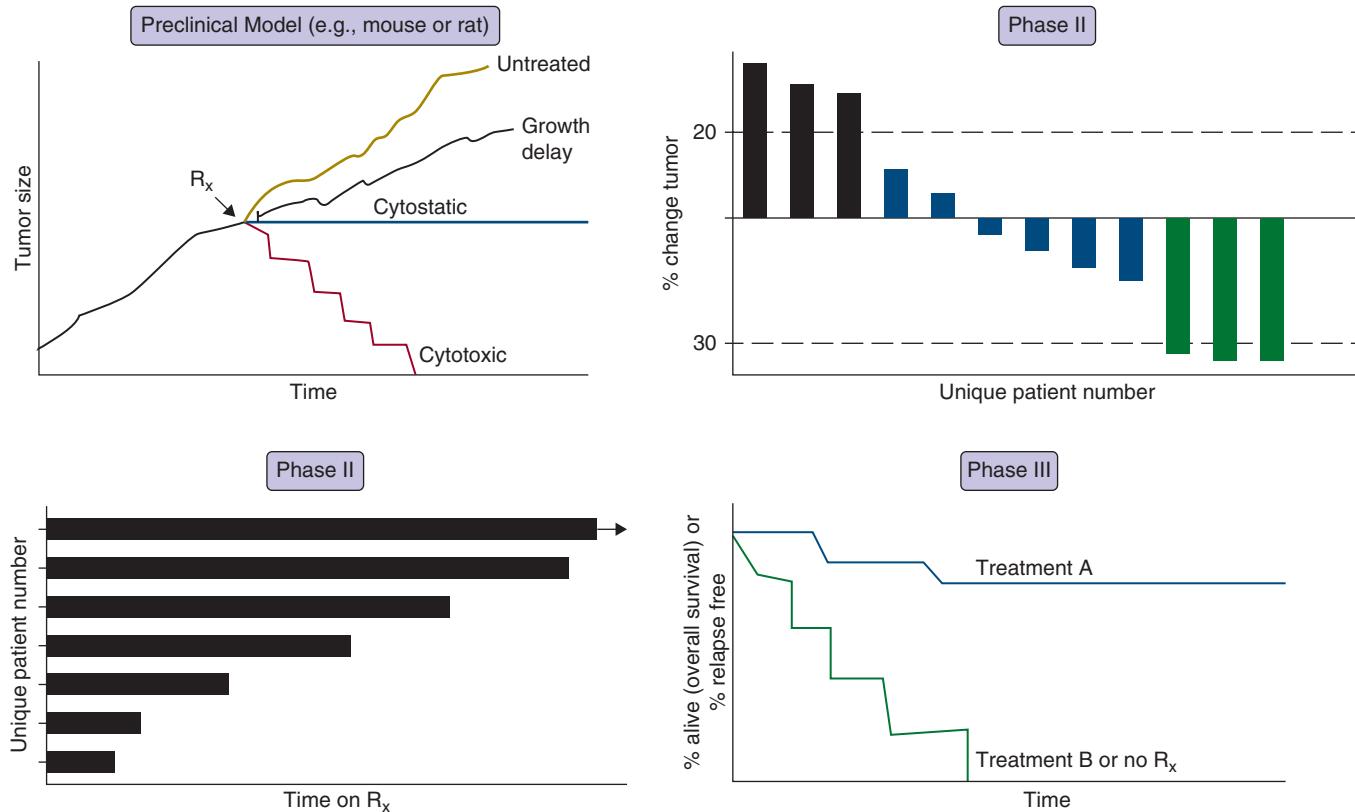


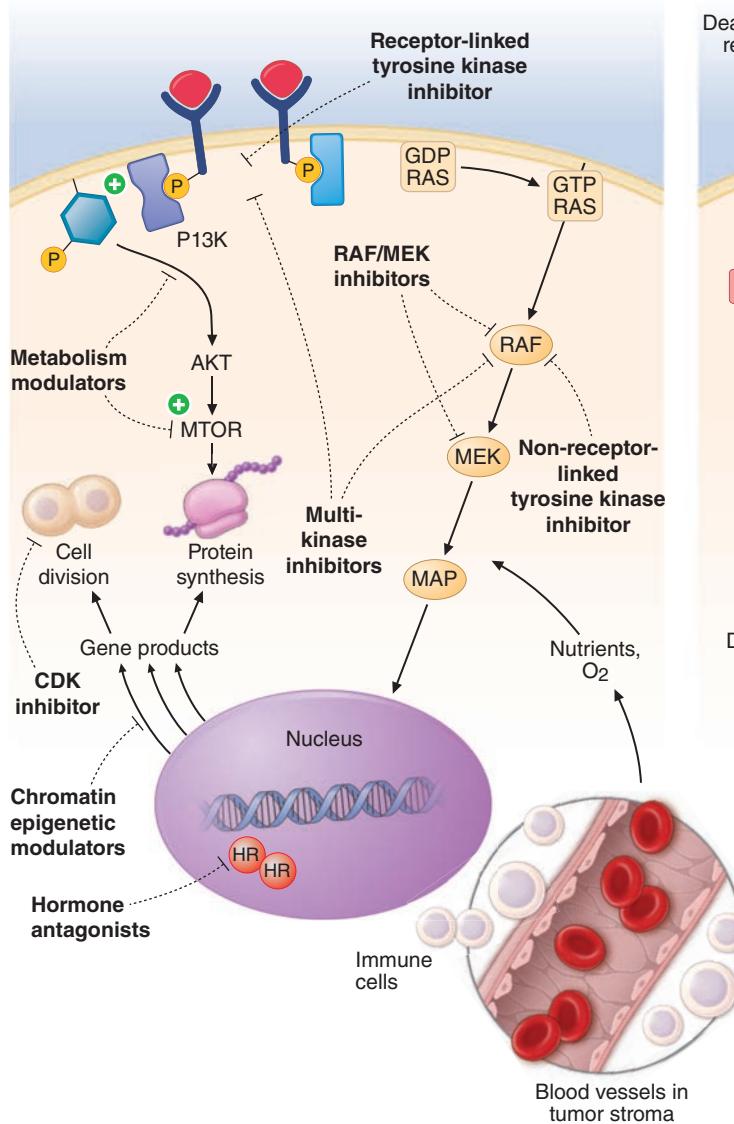
FIGURE 73-2 Steps in cancer drug discovery and development. Preclinical activity (top left) in animal models of cancers may be used as evidence to support the entry of the drug candidate into phase 1 trials in humans to define a correct dose and observe any clinical antitumor effect. The drug may then be advanced to phase 2 trials directed against specific cancer types, with rigorous quantitation of antitumor effects. Waterfall plots are a standard representation of how patients’ tumor sizes change in relation to treatment, with predefined cutoffs defining progression of disease (20% increase in size) or partial response (30% decrease in size) serving as benchmarks of potential valuable effect (top right). Swimmer plots (bottom left) allow the delineation of patients with especially long (or short) times on treatment even without response, another basis in the former case for potential perceived clinical benefit of the treatment. Kaplan-Meier plots (bottom right) of survival indices in phase 3 comparative trials may allow definition of superiority, inferiority, or no difference of treatment effect compared to standard or no treatment.

originating organ site of disease but across all organ types possessing certain molecular or biologic features.

Useful cancer drug treatment strategies using conventional chemotherapy agents, targeted agents, hormonal treatments, or biologicals all have one of two valuable outcomes. They can induce cancer cell death, resulting in tumor shrinkage with corresponding clinical benefit evidenced by improvement in patient survival, or increase in time until the disease progresses. Another potential outcome is induction

of cancer cell differentiation or dormancy with loss of tumor cell replicative potential and reacquisition of phenotypic properties resembling normal cells. Interaction of a chemotherapeutic drug with its target induces a “cascade” of further signaling steps. These signals ultimately lead to cell death by triggering an “execution phase,” where proteases, nucleases, and endogenous regulators of the cell death pathway are activated (Fig. 73-3), or differentiation by alteration of cancer genome function.

Tumor cell growth pathways



Tumor cell death pathways

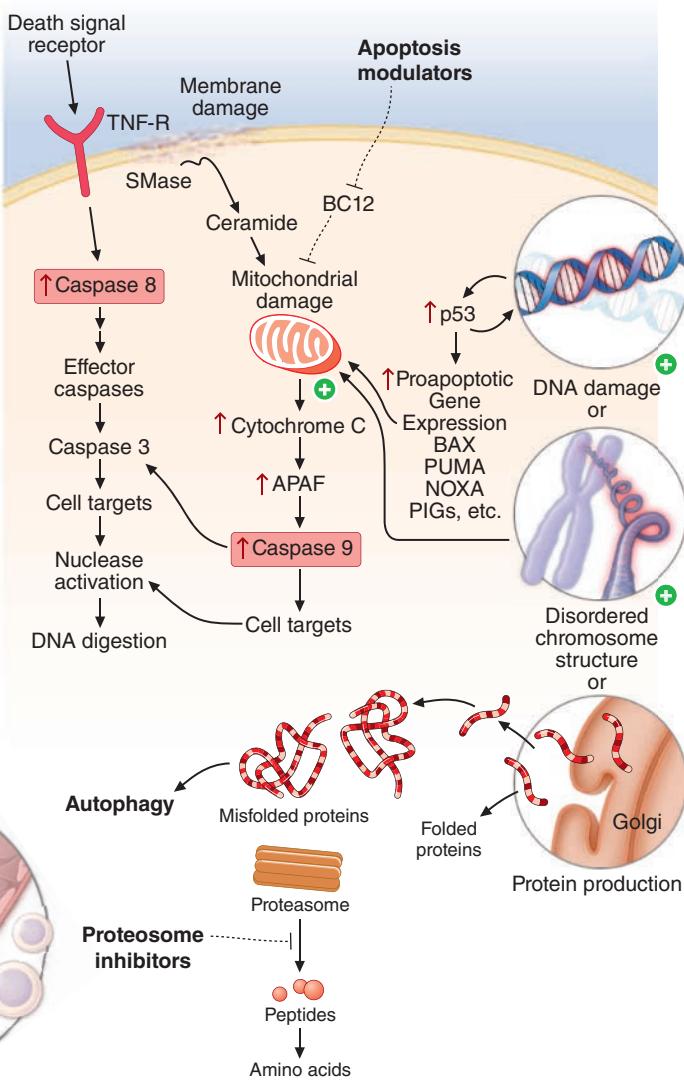


FIGURE 73-3 Tumor growth and death pathways affected by targeted and cytotoxic agents. After a growth factor binds to its receptor (left side of figure), in the most commonly activated cell proliferation pathway, increased tyrosine kinase activity occurs, either by autophosphorylation of receptor-linked kinases or through recruitment of non-receptor-linked tyrosine kinases, which may also be active constitutively, without requiring a growth factor. This leads to docking of “adaptor” proteins to the phosphorylated tyrosines. One important pathway activated occurs after exchange of GDP for GTP in the RAS family of protooncogene products. GTP-RAS activates the RAF kinase, leading to a phosphorylation cascade of MEK and MAP kinases that ultimately alters gene function to produce transcripts that activate cell cycle progression through cyclin-dependent kinases (CDKs). Another route to gene activation utilizes hormone receptors (HRs) interacting with tissue-specific growth regulators such as steroid hormones to alter gene function leading to cell cycle activation. Tyrosine phosphorylation can lead to activation of phosphatidylinositol-3-kinase (PI3K) to produce the phosphorylated lipid phosphatidylinositol-3-phosphate, which activates the AKT kinase to act downstream on the mammalian target of rapamycin kinase (mTOR), directly increasing translation of key mRNAs for gene products regulating cell growth. Cytotoxic agents cause cell death (right side of figure) through apoptosis and/or induction of autophagy. Apoptosis is also stimulated by interruption of growth factor (GF) cytokine death signals (e.g., tumor necrosis factor receptor [TNF-R]), which activate “upstream” cysteine aspartyl proteases (caspases) to directly digest cytoplasmic and nuclear proteins, resulting in activation of “downstream” caspases; these activate nucleases to cause DNA fragmentation, a hallmark of apoptosis. Chemotherapy agents that create lesions in DNA or alter mitotic spindle function activate gene function to alter mitochondrial integrity. The antiapoptotic protein BCL2 attenuates mitochondrial toxicity, whereas proapoptotic gene products such as BAX, PUMA, etc., antagonize the action of BCL2. Damaged mitochondria release cytochrome C and apoptosis-activating factor (APAF), which activate caspase 9 to cause DNA fragmentation. In addition, membrane damage with activation of sphingomyelinases results in the production of ceramides that can cause direct damage to mitochondria. Protein translation is followed by a folding process in the Golgi apparatus. Misfolded proteins are processed through the proteasome for protease digestion and recycling of amino acids. Disruption of this process can contribute to autophagy, where the cell starves for critical nutrients, or itself induce apoptosis through a distinct pathway activated by misfolded protein accumulation. Antiangiogenic agents and immune therapies work in the tumor stroma (lower left) on supporting elements including blood vessels and host inflammatory cells.

Targeted agents differ from chemotherapy agents in that they do not indiscriminately cause macromolecular lesions but regulate the action of macromolecules to whose function tumors have been described as “addicted” in the sense that without the pathway’s continued action, the tumor cell cannot survive. In this way, targeted agents directed at such “oncogenic driver” molecules may alter the threshold for tumors driven by these molecules to undergo cell death.

Strategies in Systemic Cancer Management The past 30 years have witnessed a marked evolution in the systemic treatment of cancer not amenable to cure by locally applied treatments. Nonspecific cytotoxic agents of limited efficacy for most patients but highly active and curative in a minority disease types have been joined by targeted and biologic therapies. **Table 73-3, A** lists those tumors considered curable by conventionally available chemotherapeutic agents even when disseminated or metastatic. If a tumor is truly localized to a single site, consideration of surgery or primary radiation therapy should be given as well. Chemotherapy may be used as part of multimodality approaches to offer primary treatment to a clinically localized tumor (**Table 73-3, B**). Chemotherapy can be administered as an *adjuvant*, i.e., in addition to surgery or radiation (**Table 73-3, C**), even after all clinically apparent disease has been removed. This use of chemotherapy has curative potential in, e.g., lung, breast, and colorectal

neoplasms, as it eliminates clinically unapparent tumor that may have already disseminated. *Neoadjuvant* chemotherapy refers to administration of chemotherapy before any surgery or radiation to a local tumor in an effort to enhance the effect of subsequent local treatment.

Chemotherapy is routinely used in doses that produce reversible acute side effects, primarily consisting of transient myelosuppression with or without gastrointestinal toxicity (usually nausea). “High-dose” chemotherapy regimens can produce markedly increased therapeutic effect, although at the cost of potentially life-threatening complications that require intensive support, usually in the form of hematopoietic stem cell support from the patient (*autologous*) or from donors matched for histocompatibility loci (*allogeneic*), or pharmacologic “rescue” strategies to block the effect of the high-dose chemotherapy on normal tissues. High-dose regimens have curative potential in defined clinical settings (**Table 73-3, D**).

If cure is not possible, chemotherapy may be undertaken with the goal of palliating the tumor’s effect on the host (**Table 73-3, E**). In this usage, value is perceived by the demonstration of improved symptom relief, progression-free survival, or overall survival. The data result from a clinical research protocol used as a basis for FDA approval for commercial use of the agent. Patients treated with palliative intent should be aware of their diagnosis and the limitations of the proposed treatments, have access to supportive care, and have suitable “performance status,” according to assessment algorithms such as the one developed by Karnofsky (see **Table 69-4**) or by the Eastern Cooperative Oncology Group (ECOG) (see **Table 69-5**). ECOG performance status (PS) 0 patients are without symptoms; PS1 patients are ambulatory but restricted in strenuous physical activity; PS2 patients are ambulatory and active 50% or more of the time but unable to work; PS3 patients are capable of limited self-care but are active <50% of the time; and PS4 patients are totally confined to bed or chair and incapable of self-care. Only PS0, PS1, and PS2 patients are generally considered suitable for palliative (noncurative) treatment. If there is curative potential, even poor-PS patients may be treated (especially if their symptoms are directly related to a cancer that may respond to treatment), but their prognosis is usually inferior to that of good-PS patients treated with similar regimens. Assessment of physiologic reserve through use of the geriatric assessment tool can be helpful, but no measure of comorbidities or physiologic reserve is considered standard.

The turn of the millennium marked the arrival of alternative strategies for cancer treatment. Prominent among these are *cancer biologic therapy*, which harnesses the use of immune system-derived reagents or strategies, and *cancer targeted therapies*, which are directed at specific molecular targets differentially expressed in malignant as opposed to normal tissues.

CANCER BIOLOGIC THERAPY

Figure 73-4 presents the landscape of cancer biologic therapy agents and actions. The goal of biologic therapy is to manipulate the host-tumor interaction in favor of the host, potentially at an optimum biologic dose that might be different than MTD. As a class, biologic therapies may be distinguished from cytotoxic and molecularly targeted agents in that biologic therapies require activity (e.g., antigen expression or internalization) on the part of the tumor cell or on the part of the host (e.g., T-cell engagement or cytokine elaboration) to allow therapeutic effect.

Antibody-Mediated Therapeutic Approaches Figure 73-4 illustrates current antibody-based strategies in cancer treatment. The ability to grow very large quantities of high-affinity monoclonal antibodies directed at specific tumor antigens produced by animals allows the grafting of animal-derived antigen-combining sequences into human immunoglobulin genes (chimerized or humanized products) or derived de novo from mice bearing human immunoglobulin gene loci. Four general strategies have emerged using antibodies. *Antitumor cell antibodies* target tumor cells directly to inhibit intracellular functions or attract immune or stromal cells. *Bispecific tumor engaging (BiTe) antibodies* directly bind to a tumor cell and to a host immune cell. *Immunoregulatory antibodies* target antigens expressed on host

TABLE 73-3 Clinical Impact on Cancers with Cytotoxic Chemotherapy

A. Advanced Cancers with Possible Cure	D. Cancers Possibly Cured with High-Dose Chemotherapy with Stem Cell Support
Acute lymphoid and acute myeloid leukemia (pediatric/adult)	Relapsed leukemias, lymphoid and myeloid
Hodgkin’s disease (pediatric/adult)	Relapsed lymphomas, Hodgkin’s and non-Hodgkin’s
Lymphomas—certain types (pediatric/adult)	Chronic myeloid leukemia
Germ cell neoplasms	Multiple myeloma
Embryonal carcinoma	E. Cancers Responsive with Useful Palliation, But Not Cure, by Chemotherapy
Teratocarcinoma	Bladder carcinoma
Seminoma or dysgerminoma	Chronic myeloid leukemia
Choriocarcinoma	Hairy cell leukemia
Gestational trophoblastic neoplasia	Chronic lymphocytic leukemia
Pediatric neoplasms	Lymphoma—certain types
Wilms’ tumor	Multiple myeloma
Embryonal rhabdomyosarcoma	Gastric carcinoma
Ewing’s sarcoma	Cervix carcinoma
Peripheral neuroepithelioma	Endometrial carcinoma
Neuroblastoma	Soft tissue sarcoma
Small-cell lung carcinoma	Head and neck cancer
Ovarian carcinoma	Adrenocortical carcinoma
B. Advanced Cancers Possibly Cured by Chemotherapy and Radiation	Islet cell neoplasms
Squamous carcinoma (head and neck)	Breast carcinoma
Squamous carcinoma (anus)	Colorectal carcinoma
Breast carcinoma	Renal carcinoma
Carcinoma of the uterine cervix	F. Tumors Poorly Responsive in Advanced Stages to Chemotherapy
Non-small-cell lung carcinoma (stage III)	Pancreatic carcinoma
Small-cell lung carcinoma	Biliary tract neoplasms
C. Cancers Possibly Cured with Chemotherapy as Adjuvant to Surgery	Thyroid carcinoma
Breast carcinoma	Carcinoma of the vulva
Colorectal carcinoma ^a	Non-small-cell lung carcinoma
Osteogenic sarcoma	Prostate carcinoma
Soft tissue sarcoma	Melanoma (subsets)
	Hepatocellular carcinoma
	Salivary gland cancer

^aRectum also receives radiation therapy.

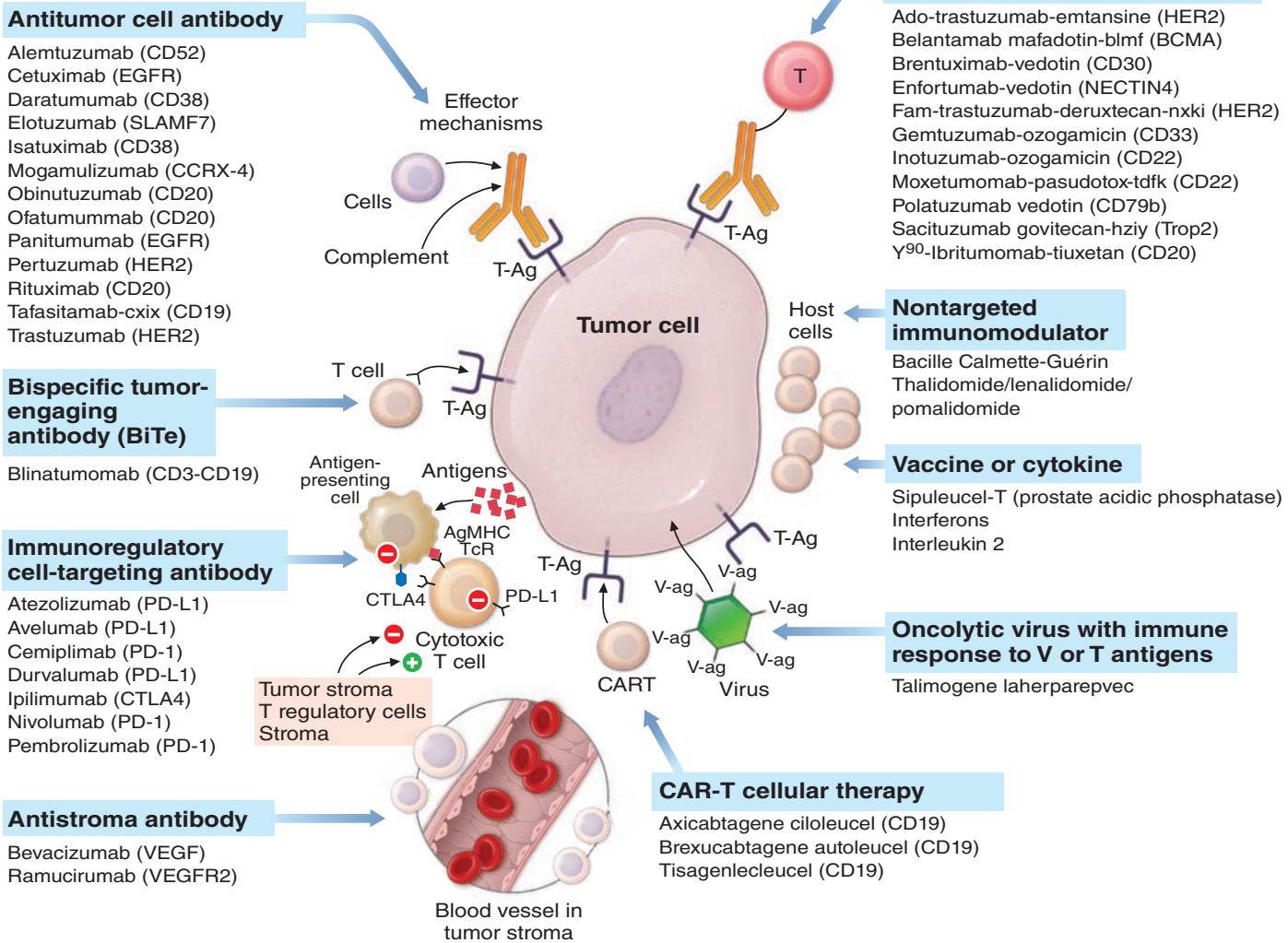


FIGURE 73-4 Immunologic treatments for cancer. Anti-tumor cell antibodies targeting antigens (T-Ag) expressed on tumor cells and indicated in parentheses for each antibody or antibody-derived construct (*upper left*) can either directly interfere with tumor cell function by, e.g., inhibiting growth-promoting pathways, or recruit host immune effector cells actively (especially through *bispecific tumor-engaging* [BiTe] strategies), Fc receptors, or cytotoxic mechanisms such as complement. Proceeding clockwise in the figure, *antibody conjugates* can also be engineered to deliver cytotoxic drugs, bacteria-derived toxins, or radioisotopes (T) to T-Ags (*upper right*). Relatively *nonspecific immunomodulators* include vaccines instilled directly into the tumor stroma, agents such as the “imids” that alter tumor and stromal cell cytokine production, and cytokines such as interferon or interleukin 2 (IL-2), which can affect tumor-infiltrating lymphocyte function or have direct antitumor effects. *Vaccines* targeting tumor cell antigens or live attenuated *oncolytic viruses* injected into tumors can cause tumor cell lysis with induction of a prominent host antitumor immune response to virus Ags and T-Ags. In the *left lower* portion of the figure, strategies to target stromal and immune cells include derivation of autologous T cells that are then infected with a lentivirus or other construct that targets antigens (T-Ags) expressed on tumor cells (*chimeric antigen receptor* [CAR] T cells), with the targeted antigen in parentheses. Alternatively, endogenous T cells can be activated by *immunomodulatory cell targeting antibodies*. Specifically, tumor cell-derived antigens are taken up by antigen-presenting cells (APCs), also in the stroma. Antigens are processed by the APCs to peptides presented by the major histocompatibility complex (MHC) to T-cell antigen receptors (TcRs), thus providing a positive (+) activation signal for the cytotoxic tumor cells to kill tumor cells bearing that antigen. Negative (-) signals inhibiting cytotoxic T-cell action include the CTLA4 receptor (on T cells), interacting with the B7 family of negative regulatory signals from APCs, and the PD receptor (on T cells), interacting with the PD ligand-1 (PD-L1) (-) signal coming from tumor cells expressing the PD-L1. Strategies that inhibit CTLA4 and PD-1 function are a means of stimulating cytotoxic T-cell activity to kill tumor cells. *Tumor stroma-directed antibodies* cause anti-vascular endothelial cell growth factor (VEGF)-mediated antiangiogenic and tumor interstitial pressure-modulating strategies.

immune cells to boost the host's immune response to the tumor. Finally, *antibody conjugates* link the antibody to drugs, toxins, or radioisotopes to target these “warheads” for delivery to the tumor. These will be considered with cytotoxic agents. *Stroma-directed antibodies* are currently available against tumor supporting vasculature.

ANTI-TUMOR CELL ANTIBODIES (FIG. 73-4) Humanized antibodies against the CD20 molecule expressed on B-cell lymphomas (rituximab and ofatumumab) are exemplary of antibodies that affect both signaling events driving lymphomagenesis as well as activating immune responses against B-cell neoplasms. They are used as single agents and in combination with chemotherapy and radiation in the treatment of B-cell neoplasms. Obinutuzumab is an antibody with altered

glycosylation that enhances its ability to activate killer cells; it is also directed against CD20 and is of value in chronic lymphocytic leukemia.

Unintended side effects of any antibody include infusion-related hypersensitivity reactions, usually limited to the first infusion, which can be managed with glucocorticoid and/or antihistamine prophylaxis, and prolonged infusion strategies.

Anti-B-cell-directed antibodies can have the unintended effect of exacerbating immunosuppression with the emergence of increased opportunistic infections. Reactivation of latent infections may also occur; an assessment of a patient's hepatitis B and C status is conventionally done before treatment. Concomitant use of antivirals directed against hepatitis may be indicated. Patients with HIV and lymphoma need antivirals optimized to minimize interaction with

anti-lymphoma treatments; consultation with infectious disease specialists is warranted. Anti-tumor cell antibodies also include approaches to activate complement and are exemplified by alemtuzumab directed against CD52; it is active in chronic lymphoid leukemia and T-cell malignancies. Tumor lysis syndrome prophylaxis may be warranted.

Epidermal growth factor receptor (EGFR)-directed antibodies (e.g., cetuximab and panitumumab) have activity in colorectal cancer refractory to chemotherapy, particularly when used to augment the activity of an additional chemotherapy program, and in the primary treatment of head and neck cancers treated with radiation therapy. Direct effects on the tumor may mediate an antiproliferative effect as well as stimulate the participation of host mechanisms involving immune cell or complement-mediated response to tumor cell-bound antibody. Anti-EGFR antibodies can cause an acneiform rash requiring topical antibiotic and glucocorticoid cream treatment; photosensitivity also occurs.

The HER2/neu receptor overexpressed on epithelial cancers, especially breast and certain gastrointestinal cancers, was initially targeted by trastuzumab, with activity in potentiating the action of chemotherapy in breast cancer as well as evidence of single-agent activity. Trastuzumab appears to interrupt intracellular signals derived from HER2/neu and to stimulate anti-tumor cell immune mechanisms. The anti-HER2 antibody pertuzumab, specifically targeting the domain of HER2/neu responsible for dimerization with other HER2 family members, is more specifically directed against HER2 signaling function and augments the action of trastuzumab. Both trastuzumab and pertuzumab can damage cardiac function, particularly in patients with prior exposure to anthracyclines, and left ventricular function should be checked pre-treatment and monitored during treatment.

The BiTe antibody blinatumomab was constructed to have an anti-CD19 antigen-combining site directed at a cancer cell as one valency with an anti-CD3 binding site as the other. This antibody can bring T cells (with their anti-CD3 activity) close to neoplastic B cells bearing the CD19 determinant. Blinatumomab is active in B-cell neoplasms such as acute lymphocytic leukemia. Unique toxicities include cytokine release syndrome (fever, hypotension, tachycardia) and neurologic deterioration manifest initially by deterioration in handwriting accuracy, which can proceed to more florid cortical dysfunction, suggesting a need for dose pausing and/or glucocorticoid use.

STROMA-DIRECTED ANTIBODIES (FIG. 73-4) The anti-vascular endothelial growth factor (VEGF) antibody bevacizumab shows some evidence of value in renal cancers, where activation of VEGF signaling occurs as part of disabled hypoxia-induced signaling in the tumor cells. When combined with chemotherapeutic agents, it may increase responses in colorectal and nonsquamous lung cancers. The mechanism for this effect may relate to improved delivery and tumor uptake of the active chemotherapeutic agent, owing to decreased tumor interstitial pressure. VEGF was originally isolated as a “tumor permeability factor” causing leakiness of tumor blood vessels. When used in gliomas, it may, by decreasing vascular permeability, allow replacement of steroids to decrease intracranial pressure. Bevacizumab is directed against VEGF, which is normally a secreted product and not attached to tumor cells. Bevacizumab has a number of side effects including hypertension, thrombosis, proteinuria, hemorrhage, and gastrointestinal perforations with or without prior surgeries; these adverse events also occur with small-molecule drugs modulating VEGFR function.

IMMUNOREGULATORY ANTIBODIES (FIG. 73-4) Purely immunoregulatory antibodies stimulate immune responses to mediate tumor-directed cytotoxicity. An understanding of the tumor-host interface has revealed that cytotoxic tumor-directed T cells are frequently inhibited by ligands upregulated in the tumor cells. The programmed death ligand 1 (PD-L1; also known as B7-homolog 1) was initially recognized as inducing T cell death through a receptor present on T cells, termed the programmed death (PD) receptor, which physiologically regulates the intensity of the immune response to any antigen. The PD family of ligands and receptors also regulates macrophage function, present in tumor stroma. These actions raised the hypothesis that antibodies directed against the PD signaling axis (both anti-PD-L1 and anti-PD)

might be useful in cancer treatment by allowing reactivation of the immune response against tumors.

Ipilimumab, an antibody directed against the anti-CTLA4 (cytotoxic T lymphocyte antigen 4), which is expressed on T cells (not tumor cells), responds to signals from antigen-presenting cells (Fig. 73-4) and also downregulates the intensity of the T-cell proliferative response to antigens derived from tumor cells. Indeed, manipulation of the CTLA4 axis was the first demonstration that purely immunoregulatory antibody strategies directed at T-cell physiology could be safe and effective in the treatment of cancer. Ipilimumab, alone or in combination with PD-1-directed antibodies, is approved for treatment of metastatic melanoma and lung cancers.

Nivolumab, directed against the PD-1 receptor, or atezolizumab (anti-PD-L1) are exemplary of anti-PD-1 directed immunoregulatory antibodies, with clinical benefits in many cancers (Table 73-4). Pembrolizumab is approved for first-line treatment of metastatic non-small-cell lung cancer tumors that express the PD-L1 ligand. This development was a milestone in cancer therapeutics, replacing chemotherapy in this patient subset.

Importantly, the clinical observation that tumors most amenable to treatment with immunoregulatory antibodies were in sites (lung, skin, genitourinary) exposed to environmental carcinogens or occurred in patients with known mutations in DNA repair pathways stimulated specific research as to whether the “mutational burden” of tumors could predict value from anti-PD strategies. Results to date confirm this hypothesis and led to the first regulatory approvals for immunomodulating antibodies in a “tissue agnostic” fashion. Specifically, detection of deficiencies in a tumor DNA mismatch repair system or with evidence of increased tumor mutation burden is a specific indication for use of certain immunoregulatory agents, irrespective of the disease site of origin. The increased efficacy in the setting of higher tumor mutational burden is thought to be due to the presence of more proteins in the tumor structurally altered by mutation that can be recognized as foreign by the immune system.

Prominent autoimmune hepatic, endocrine, cutaneous, neurologic, and gastrointestinal adverse events can occur with the use of ipilimumab as well as the PD-1-directed antibodies. Emergency use of glucocorticoids may be required to attenuate severe toxicities. Although theoretically such glucocorticoid use can attenuate the antitumor effect, response rates do not appear to be compromised by their use to abrogate serious organ toxicity attributable to use of immunomodulatory antibodies. Importantly for the general internist, immunologic toxicities can occur late after exposure to the modulators of PD and CTLA4 action, even while the patient may have sustained control of tumor growth.

Nontargeted Immunomodulators (Fig. 73-4) Bacille Calmette-Guérin, a killed mycobacterial product, invokes a useful immune response when instilled locally into the bladder in the setting of preinvasive bladder cancers. The “imids” thalidomide, lenalidomide, and pomalidomide alter cytokine elaboration in the tumor microenvironment and have antiangiogenic actions. They are a cornerstone in the management of multiple myeloma. Thromboses (warranting consideration of prophylactic anticoagulation), gastrointestinal and neuropathic adverse events, and prominent teratogenicity can occur as a consequence of their use.

Cytokines Only interferon α (IFN- α) and interleukin 2 (IL-2) are in routine clinical use. IFN is not curative for any tumor but can induce partial responses in follicular lymphoma, hairy cell leukemia, chronic myeloid leukemia, melanoma, and Kaposi’s sarcoma. It produces fever, fatigue, a flulike syndrome, malaise, myelosuppression, and depression and can induce clinically significant autoimmune disease.

IL-2 exerts its antitumor effects indirectly through augmentation of immune function. Its biologic activity is to promote the growth and activity of T cells and natural killer (NK) cells. High doses of IL-2 can produce tumor regression in certain patients with metastatic melanoma and renal cell cancer. About 2–5% of patients may experience complete remissions that are durable. Patients may require blood pressure support and intensive care to manage the toxicity. However, once

TABLE 73-4 Clinical Impact of Host T Lymphocyte–Modified Cells^a or Host T Lymphocyte–Directed Immunoregulatory Antibodies^c

A. Advanced Cancers with Positive Effect (at least 25% of treated patients have stable disease or progression-free survival of ≥27 weeks or better) or Frequent or Unexpected Prolonged Responders (efficacy may be limited to CD expression-dependent or PD-1 ligand-expressing subtypes)
Acute lymphoid leukemia ^b
Adrenocortical carcinoma ^c
Breast cancer, hormone receptor negative, HER2 negative (with chemotherapy) ^c
Colorectal cancer (microsatellite instability-high [MSI-H] or mismatch repair deficient, following treatment with fluoropyrimidine, oxaliplatin, and irinotecan) ^c
Cervix, squamous carcinoma ^c
Cutaneous, squamous carcinoma ^c
Diffuse large B-cell non-Hodgkin's lymphoma, not otherwise specified ^a
Diffuse large B-cell non-Hodgkin's lymphoma, primary mediastinal subtype ^b
Endometrial carcinoma (with lenvatinib, if microsatellite instability-stable [MSI-S] or mismatch repair wild-type) ^c
Esophageal squamous carcinoma
Gastric/gastroesophageal adenocarcinoma ^c
Head and neck squamous carcinoma ^c
Hepatocellular cancer (after sorafenib) ^c
Hodgkin's disease ^c
Mantle cell lymphoma ^a
Melanoma ^c
Merkel cell carcinoma ^c
MSI-H or mismatch repair-deficient solid tumors without satisfactory alternative ^c
Mycosis fungoides ^c
Multiple myeloma ^a
Non-small-cell lung carcinoma ^c
Paraganglioma/pheochromocytoma ^c
Renal cell carcinoma ^c
Small-cell lung carcinoma ^c
Solid tumors with high tumor mutational burden (TMB) (≥ 10 mutations/megabase) that have progressed following prior therapy without satisfactory alternative treatment ^c
Thyroid carcinoma ^c
Urothelial carcinoma ^c (including bladder, ureter)
B. Advanced Cancers with Insufficient Data to Support Host T Lymphocyte or Immunoregulatory Antibodies
Acute myeloid leukemia
Anus, squamous carcinoma
Breast cancer, hormone receptor positive
Breast cancer, hormone receptor negative, HER2 positive
Biliary tract cancers (if MSI-S or mismatch repair wild-type)
Chronic lymphocytic leukemia
Chronic myeloid leukemia
Gastrointestinal neuroendocrine/islet cell carcinoma
Gloma, all grades including glioblastoma
Germ cell neoplasms
Ovarian cancer
Osteogenic sarcoma
Pancreas adenocarcinoma
Pediatric tumors (Wilms', rhabdomyosarcoma, Ewing's, neuroblastoma, osteosarcoma)
Prostate adenocarcinoma
Salivary gland carcinoma
Soft tissue sarcoma
T-cell non-Hodgkin's lymphoma (except mycosis fungoides)
Vulva, squamous carcinoma

^aChimeric antigen receptor (CAR)-modified autologous T cells in relapsed or refractory cases. ^bBoth CAR-modified autologous T cells or an immunoregulatory antibody. ^cT-cell directed immunoregulatory antibody strategies including anti-PD1 and/or anti-PD-L1 antibodies; or BiTe antibodies against a particular tumor cell antigen.

the agent is stopped, most of the toxicities reverse completely within 3–6 days.

T Cell–Mediated Therapies The strongest evidence that the immune system can exert clinically meaningful antitumor effects comes from allogeneic bone marrow transplantation. Adoptively transferred T cells from the donor expand in the tumor-bearing host, recognize the tumor as being foreign, and can mediate impressive antitumor effects (graft-versus-tumor effects). Three types of currently used cancer treatments take advantage of the ability of T cells to kill tumor cells.

- Transfer of allogeneic T cells.* This occurs in three major settings: in allogeneic bone marrow transplantation; as purified lymphocyte transfusions following bone marrow recovery after allogeneic bone marrow transplantation; and as pure lymphocyte transfusions following immunosuppressive (nonmyeloablative) therapy (also called reduced-intensity or minitransplants). In each of these settings, the effector cells are donor T cells that recognize the tumor as being foreign, probably through minor histocompatibility differences. The main risk of such therapy is the development of graft-versus-host disease because of the minimal difference between the cancer and the normal host cells. This approach has been highly effective in certain hematologic cancers refractory to chemotherapeutic strategies.
- Transfer of autologous T cells.* In this approach, the patient's own T cells are removed from the tumor-bearing host, manipulated in several ways *in vitro*, and given back to the patient. Tumor antigen-specific T cells can be developed after retroviral transduction of the desired T-cell antigen receptor and expanded to large numbers over many weeks *ex vivo* before administration. These chimeric antigen receptor (CAR) T cells (Fig. 73-4) have evidence of sustained value in patients with refractory hematopoietic neoplasms such as diffuse large B-cell lymphoma and mantle cell lymphoma. Prominent adverse effects include cytokine release (fever, tachycardia, hypotension) and neurologic manifestations. Clinical investigations are seeking to develop solid-tumor antigen-directed CAR strategies, as well as to utilize different immune cell populations such as NK cells to deliver the antigen receptor construct in ways that may allow "off-the-shelf" products not requiring manipulation and purification of patients' autologous cells.

A second autologous T-cell strategy uses activation of the patient's T cells to polyclonal stimulators such as anti-CD3 and anti-CD28 after a short period *ex vivo* and then amplification in the host after transfer by stimulation with IL-2. Short periods removed from the patient permit the cells to overcome the tumor-induced T-cell defects, and such cells traffic and home to sites of disease better than cells that have been in culture for many weeks.

- Tumor vaccines aimed at boosting T-cell immunity.* Two types of vaccine approaches are currently approved. Purified autologous antigen-presenting cells can be pulsed with tumor, its membranes, or particular tumor antigens and delivered as a vaccine. Vaccine adjuvants such as granulocyte-macrophage colony-stimulating factor (GM-CSF) may be co-administered. One such vaccine, sipuleucel-T (Fig. 73-4), is approved for use in patients with hormone-independent prostate cancer. In this approach, the patient undergoes leukapheresis, wherein mononuclear cells (that include antigen-presenting cells) are removed from the patient's blood. The cells are pulsed in a laboratory with an antigenic fusion protein comprising a protein frequently expressed by prostate cancer cells, prostate acid phosphatase, fused to GM-CSF, and matured to increase their capacity to present the antigen to immune effector cells. The cells are then returned to the patient in a well-tolerated treatment. Although no objective tumor response was documented in clinical trials, median survival was increased by about 4 months.

Another important vaccine strategy is directed at infectious agents whose action ultimately is tied to the development of human cancer. Hepatitis B vaccine in an epidemiologic sense prevents hepatocellular carcinoma, and a tetravalent human papillomavirus vaccine prevents infection by virus types currently accounting for 70% of cervical cancer. Unfortunately, these vaccines are ineffective at treating patients who have developed a virus-induced cancer.

Oncolytic Viruses (Fig. 73-4) Laboratory studies in animals have utilized viruses to destroy tumors because tumor cells lack endogenous host mechanisms, e.g., IFN elaboration or recognition strategies of viral nucleic acids, that limit virus spread. Viral infection of tumors also can stimulate a prominent host response to viral and tumor cell antigens, leading to immune effects against local tumor cells. Talimogene laherparepvec is a clinically approved attenuated herpes virus that acts to stimulate immune responses when instilled locally into melanoma deposits. Systemic effects are minimal in this application. This general strategy is being considered particularly in tumors not amenable to useful effects of currently approved immunoregulatory antibodies or in conjunction with immunoregulatory antibodies.

CANCER CYTOTOXIC THERAPY

Table 73-5 lists commonly used cytotoxic cancer chemotherapy agents and pertinent clinical aspects of their use, with particular reference to adverse effects that might be encountered by the generalist in the care of patients. The drugs were initially discovered through screening of chemicals and natural product extracts to define evidence of antitumor activity in animals or were designed with knowledge of biochemical pathways affecting nucleic acid synthesis. They may be usefully grouped into two general categories: those affecting DNA and those affecting microtubules.

As illustrated in Fig. 73-3, disruption of DNA or microtubule integrity is a major trigger of cellular apoptosis pathways. An additional factor in drug effect stems from recent observations that tumor cells have increased tolerance of specific types of DNA damage owing to defects in DNA repair pathways. This state is thought to facilitate the survival of the neoplastic clone as it experiences DNA mutations during the course of carcinogenesis. DNA-directed cytotoxic agents can interact with certain DNA repair mutations in a “synthetic lethal” fashion: the DNA repair mutation enhances lethality of the chemotherapy agent. Examples of a potential “synthetic lethal effect” will be pointed out in relation to clinical applications below.

DNA-Interactive Agents DNA replication occurs during the synthesis or S-phase of the cell cycle, with chromosome segregation of the replicated DNA in the M, or mitosis, phase. The G₁ and G₂ “gap phases” precede S and M, respectively. Chemotherapeutic agents have been divided into “phase-nonspecific” agents, which can act in any phase of the cell cycle, and “phase-specific” agents, which require the cell to be at a particular cell cycle phase to cause greatest effect.

Alkylating agents (Table 73-5) as a class are cell cycle phase-nonspecific agents. They break down, either spontaneously or after normal organ or tumor cell metabolism, to reactive intermediates that covalently modify bases in DNA. This leads to cross-linkage of DNA strands or the appearance of breaks in DNA as a result of repair efforts. Damaged DNA cannot complete normal cell division; in addition, it activates apoptosis. Alkylating agents share similar toxicities: myelosuppression, alopecia, gonadal dysfunction, mucositis, and pulmonary fibrosis. They also share the capacity to cause “second” neoplasms, particularly leukemia, years after use, particularly when used in low doses for protracted periods.

Cyclophosphamide is inactive unless metabolized by the liver to 4-hydroxy-cyclophosphamide, which decomposes into an alkylating species, as well as to chloroacetaldehyde and acrolein. The latter causes chemical cystitis; therefore, excellent hydration must be maintained while using cyclophosphamide. If severe, the cystitis may be attenuated or prevented altogether (if expected from the dose of cyclophosphamide to be used) by mesna (2-mercaptopethanesulfonate). Liver disease impairs cyclophosphamide activation. Sporadic interstitial pneumonitis leading to pulmonary fibrosis can accompany the use of cyclophosphamide, and high doses used in conditioning regimens for bone marrow transplant can cause cardiac dysfunction. Ifosfamide is a cyclophosphamide analogue also activated in the liver, but more slowly, and it requires co-administration of mesna to prevent bladder injury. CNS effects, including somnolence, confusion, and psychosis, can follow ifosfamide use; the incidence appears related to low body surface area or decreased creatinine clearance.

Several alkylating agents are less commonly used. Bendamustine has activity in chronic lymphocytic leukemia and certain lymphomas. Busulfan can cause profound myelosuppression, alopecia, and pulmonary toxicity but is relatively “lymphocyte sparing.” It is used in transplant preparation regimens. Melphalan shows variable oral bioavailability and undergoes extensive binding to albumin and α_1 -acidic glycoprotein. Mucositis appears more prominently; however, it has prominent activity in multiple myeloma.

Nitrosoureas break down to carbamylating species that not only cause a distinct pattern of DNA base pair-directed reactivity but also can covalently modify proteins. They share the feature of causing relatively delayed bone marrow toxicity, which can be cumulative and long-lasting. Procarbazine is metabolized in the liver and possibly in tumor cells to yield a variety of free radical and alkylating species. In addition to myelosuppression, it causes hypnotic and other CNS effects, including vivid nightmares. It can cause a disulfiram-like syndrome on ingestion of ethanol. Dacarbazine (DTIC) is activated in the liver to yield the highly reactive methyl diazonium cation. It causes only modest myelosuppression 21–25 days after a dose but causes prominent nausea on day 1. Temozolomide is structurally related to dacarbazine but is activated by nonenzymatic hydrolysis in tumors and is bioavailable orally. Brain tumors with alkylguanine alkyl transferase deficiency are selectively susceptible to temozolomide, which alkylates the O⁶ position of guanine.

Cisplatin was discovered fortuitously by observing that bacteria present in electrolysis solutions with platinum electrodes could not divide. Only the *cis* diamine configuration is active as an antitumor agent. In tumor cells, a chloride is lost from each position. The resulting positively charged species is an efficient DNA interactor, forming Pt-based cross-links. Cisplatin is administered with abundant hydration, including forced diuresis with mannitol to prevent kidney damage; even with the use of hydration, gradual decrease in kidney function is common, along with noteworthy anemia. Hypomagnesemia frequently attends cisplatin use and can lead to hypocalcemia and tetany. Other common toxicities include neurotoxicity with stocking-and-glove sensorimotor neuropathy. Hearing loss occurs in 50% of patients treated with conventional doses. Cisplatin is intensely emetogenic, requiring prophylactic antiemetics. Myelosuppression is less evident than with other alkylating agents. Chronic vascular toxicity (Raynaud’s phenomenon, coronary artery disease) is a more unusual toxicity. Carboplatin displays less nephro-, oto-, and neurotoxicity. However, myelosuppression is more frequent, and because the drug is exclusively cleared through the kidney, adjustment of dose for creatinine clearance must be accomplished through use of various dosing nomograms. Oxaliplatin is a platinum analogue with noteworthy activity in colon cancers refractory to other treatments. It is prominently neurotoxic.

Trabectedin binds to DNA through the “DNA minor groove” with covalent interaction with the N2 position of certain guanines. Uniquely among the DNA interactors, this can lead to the disruption of the selective FUS-CHOP transcription factor action, important in the pathogenesis of certain liposarcomas. Transient altered liver function can occur, as well as cytopenias. Lurbinectedin is an analogue of trabectedin and also alters RNA polymerase function after binding to the minor groove of DNA, but has a distinct pharmacologic profile.

Antitumor Antibiotics and Topoisomerase Poisons Antitumor antibiotics are substances produced by bacteria that provide a chemical defense against hostile microorganisms. They bind to DNA directly and can frequently undergo electron transfer reactions to generate free radicals in close proximity to DNA, leading to DNA damage in the form of single-strand breaks or cross-links. Topoisomerase poisons include natural products or semisynthetic derivatives that modify enzymes that allow DNA to unwind during replication or transcription. These include topoisomerase I, which creates single-strand breaks that then rejoin following the passage of the other DNA strand through the break. Topoisomerase II creates double-strand breaks through which another segment of DNA duplex passes before rejoining. Owing to the role of topoisomerase I in the replication fork, topoisomerase I poisons cause lethality if the topoisomerase I-induced lesions occur in S-phase.

TABLE 73-5 Cytotoxic Chemotherapy Agents^a

DRUG	TOXICITY	INTERACTIONS, ISSUES
Direct DNA-Interacting Agents^a		
Alkylator		
Bendamustine	Contraindicated with prior sensitivity to polyethylene glycol 400, propylene glycol, or monothioglycerol; cytopenias, infections, cutaneous eruptions, hepatotoxicity	Monitor for tumor lysis syndrome, extravasation, anaphylaxis, and infusion reactions
Carboplatin	Marrow: platelets > WBCs; nausea, renal (high dose)	Reduce dose according to CrCl: to AUC of 5–7 mg/mL per min [AUC = dose/(CrCl + 25)]
Carmustine (BCNU)	Myeloid (delayed nadir), GI, liver (high dose), renal	Pulmonary toxicity especially after >1400 mg/m ² cumulative dose; can be delayed in appearance
Cisplatin	Nausea, neuropathy, auditory, marrow: platelets > WBCs; renal, ↓Mg ²⁺ , ↓Ca ²⁺	Maintain high urine flow; osmotic diuresis, monitor intake/output, K ⁺ , Mg ²⁺ ; emetogenic—prophylaxis needed; full dose if CrCl >60 mL/min and tolerate fluid push
Chlorambucil	Common alkylator ^b	
Cyclophosphamide	Marrow (relative platelet sparing), cystitis, common alkylator ^b , fullike symptoms, cardiac (high dose)	Liver metabolism required to activate to phosphoramide mustard + acrolein; mesna protects against “high-dose” bladder damage
Dacarbazine (DTIC)	Myelosuppressive, cystitis, neurologic, metabolic acidosis	Metabolic activation
Ifosfamide	Marrow, bladder, CNS	Analog of cyclophosphamide, must use concomitant mesna
Lomustine (CCNU)	Marrow (delayed nadir)	
Lurbinectedin	Marrow, hepatotoxicity, nausea, vomiting	CYP3A4
Melphalan	Marrow (delayed nadir), GI (high dose)	Decreased renal function delays clearance
Oxaliplatin	Nausea, anemia	Acute reversible neurotoxicity; chronic sensory neurotoxicity cumulative with dose; reversible laryngopharyngeal spasm
Procarbazine	Marrow, nausea, neurologic, common alkylator ^b	Liver and tissue metabolism required, disulfiram-like effect with alcohol, acts as MAOI. HBP after tyrosinase-rich foods
Temozolamide	Nausea, vomiting, headache, fatigue, constipation	Myelosuppression
Trabectedin	Neutropenia, risk of fever; thrombocytopenia; rhabdomyolysis, reversible hepatic toxicity but dose reduce with liver impairment	Unusual capillary leak risk; CYP3A4
Antitumor Antibiotics and Topoisomerase Poisons		
Bleomycin	Pulmonary, skin, Raynaud's, hypersensitivity	Monitor DLCO during treatment (inactivate by bleomycin hydrolase; decreased in lung/skin); O ₂ enhances pulmonary toxicity; cisplatin-induced decrease in CrCl may increase skin/lung toxicity; reduce dose if CrCl <60 mL/min
Dactinomycin	Marrow, nausea, mucositis, vesicant, alopecia	Radiation recall
Doxorubicin and daunorubicin	Marrow, mucositis, alopecia, cardiac acute/chronic, vesicant	Heparin aggregate; coadministration increases clearance; acetaminophen, BCNU increase liver toxicity; radiation recall; dose reduce with increased bilirubin
Epirubicin	Marrow, mucositis, alopecia, cardiac acute/chronic, vesicant	Dose reduce with increased bilirubin, decreased CrCl
Etoposide (VP16-213)	Marrow (WBCs > platelet), alopecia, hypotension, hypersensitivity with rapid IV, nausea, mucositis	Hepatic metabolism and renal excretion (30%); reduce doses with liver and kidney failure; accentuate antimetabolite action
Idarubicin	Marrow, mucositis, alopecia, cardiac acute/chronic, vesicant	Dose reduce with increased bilirubin, decreased CrCl
Irinotecan	Diarrhea: “early onset” with cramping, flushing, vomiting; “late onset” after several doses; marrow, alopecia, nausea, vomiting, pulmonary	Prodrug requires enzymatic clearance to active drug SN-38; early diarrhea due to acetylcholine release, can counter with atropine; late diarrhea, use loperamide 4 mg with first stool then 2 mg q2h until 12 h without stool up to 16 mg/24 h
Mitoxantrone	Marrow, cardiac (less than doxorubicin), vesicant; blue urine, nails, and sclerae	Interacts with heparin; less alopecia, nausea than doxorubicin; radiation recall; less alopecia, nausea than doxorubicin
Topotecan	Marrow, mucositis, nausea, alopecia	Reduce dose with renal failure; rare interstitial pneumonitis
Indirectly DNA-Interacting Agents		
Antimetabolites		
Asparaginase	Decrease protein synthesis; indirect inhibition of DNA synthesis by decreased histone synthesis; clotting factors; glucose; albumin hypersensitivity; CNS; pancreatitis; hepatic	Blocks methotrexate action
Capecitabine	Diarrhea, hand-foot syndrome	Prodrug of 5-fluorouracil due to intratumoral metabolism
2-Chlorodeoxadenosine	Marrow, renal, fever	Notable use in hairy cell leukemia
Cytosine arabinoside	Marrow, mucositis, neurologic (high dose), conjunctivitis (high dose; use steroid eyedrops until 72 h after last dose), noncardiogenic pulmonary edema	Metabolized in tissues by deamination but renal excretion prominent at doses >500 mg; therefore, dose reduce in “high-dose” regimens in patients with decreased CrCl
Fludarabine phosphate	Marrow, neurologic, lung	Dose reduction with renal failure; metabolized to F-ara, converted to F-ara ATP in cells by deoxyribonucleoside kinase
5-Fluorouracil	Marrow, mucositis, neurologic, skin changes	Toxicity enhanced by leucovorin by increasing “ternary complex” with thymidylate synthase; dihydropyrimidine dehydrogenase deficiency increases toxicity; metabolism in tissue
Gemcitabine	Marrow, nausea, hepatic, fever/“flu syndrome”	Rare pulmonary/capillary leak syndrome; rare hemolytic-uremic syndrome; rare posterior reversible encephalopathy syndrome; radiosensitization

(Continued)

TABLE 73-5 Cytotoxic Chemotherapy Agents^a (Continued)

DRUG	TOXICITY	INTERACTIONS, ISSUES
Hydroxyurea	Marrow, nausea, mucositis, skin, rare renal, liver, lung, CNS	Decrease dose with renal failure; augments antimetabolite effect
6-Mercaptopurine (6-MP)	Marrow, liver, nausea	Variable bioavailability, metabolized by xanthine oxidase, decrease dose with allopurinol; increased toxicity with thiopurine methyltransferase deficiency
Methotrexate	Marrow, liver, lung, renal tubular, mucositis	Toxicity lessened by “rescue” with leucovorin; excreted in urine; decrease dose in renal failure; NSAIDs increase renal toxicity
Pemetrexed	Anemia; neutropenia	Supplement folate/B ₁₂ Caution in renal failure
Pralatrexate	Thrombocytopenia, myelosuppression, mucositis	Active in peripheral T-cell lymphoma
6-Thioguanine	Marrow, liver, nausea	Variable bioavailability; increased toxicity with thiopurine methyltransferase deficiency
Trifluridine/tipiracil	Marrow, mucositis, nausea, vomiting, unusual hand/foot	Trifluridine directly inhibits thymidylate synthase and is incorporated into DNA; tipiracil inhibits thymidine phosphorylase, which degrades trifluridine
Antimitotic Agents		
Docetaxel	Hypersensitivity to vehicle; fluid retention syndrome; marrow; dermatologic; peripheral neuropathy; nausea infrequent; some stomatitis	Premedicate with steroids, H ₁ and H ₂ blockers; may require lengthened infusions to avoid hypersensitivity
Eribulin	Marrow; peripheral neuropathy; QT prolongation	Dose modify in liver and kidney impairment
Ixabepilone	Myelosuppression, neuropathy, hypersensitivity to infusion	Premedicate with steroids, H ₁ and H ₂ blockers; may require lengthened infusions to avoid hypersensitivity; dose modification for liver impairment; CYP3A4
Nab-paclitaxel (protein bound)	Neuropathy, anemia, marrow	Dose adjust with liver dysfunction; caution with inhibitors or inducers of either CYP2C8 or CYP3A4
Paclitaxel	Hypersensitivity to vehicle; marrow; alopecia, mucositis, peripheral neuropathy, CV conduction, infrequent nausea	Premedicate with steroids, H ₁ and H ₂ blockers; hepatic clearance with dose reduction with increased bilirubin; caution with inhibitors or inducers of either CYP2C8 or CYP3A4
Vinblastine	Vesicant; marrow; peripheral neuropathy (less common but similar spectrum to other vincas); hypertension, Raynaud's, ileus/constipation (use prophylactic stool softeners)	Hepatic clearance; dose reduction for bilirubin >1.5 mg/dL
Vincristine	Vesicant, marrow (less than vinblastine), neurologic, GI; ileus/constipation (use prophylactic stool softeners); SIADH; rare CV	Hepatic clearance; dose reduction for bilirubin >1.5 mg/dL
Vinorelbine	Vesicant, marrow, allergic bronchospasm (immediate), dyspnea/cough (subacute), neuropathic (less prominent but similar spectrum to other vincas)	Hepatic clearance; dose reduction for bilirubin >1.5 mg/dL

^aAll agents in this category should be regarded as potentially fetotoxic, and use during pregnancy is either contraindicated or undertaken with clear understanding of risk of fetal harm; likewise not recommended for use during lactation. ^bCommon alkylator: alopecia, pulmonary, infertility, plus teratogenesis.

Abbreviations: ATP, adenosine triphosphate; AUC, area under the curve; CNS, central nervous system; CrCl, creatinine clearance; CV, cardiovascular; GI, gastrointestinal; CYP3A4, avoid concomitant strong CYP3A inhibitors and avoid concomitant strong CYP3A inducers; DLCO, diffusing capacity of carbon monoxide; F-ara, fludarabine; HBP, high blood pressure; MAOI, monoamine oxidase inhibitor; NSAID, nonsteroidal anti-inflammatory drug; SIADH, syndrome of inappropriate antidiuretic hormone secretion; WBC, white blood cells.

Doxorubicin intercalates into DNA, thereby altering DNA structure, replication, and topoisomerase II function. It can also undergo reduction of its quinone ring system, with reoxidation to form reactive oxygen radicals. It causes predictable myelosuppression, alopecia, nausea, and mucositis. In addition, it causes acute cardiotoxicity in the form of atrial and ventricular dysrhythmias, but these are rarely of clinical significance. In contrast, cumulative doses >550 mg/m² are associated with a 10% incidence of chronic cardiomyopathy. The incidence of cardiomyopathy appears to be related to peak serum concentration, with low-dose, frequent treatment or continuous infusions better tolerated than intermittent higher-dose exposures. Cardiotoxicity has been related to iron-catalyzed oxidation and reduction of doxorubicin in the heart. Dexrazoxane is an intracellular chelating agent that can act as a cardio-protectant. Doxorubicin's cardiotoxicity is increased when given together with trastuzumab, the anti-HER2/neu antibody. Radiation recall or interaction with concomitantly administered radiation to cause local site complications is frequent. The drug is a powerful vesicant, with necrosis of tissue apparent 4–7 days after an extravasation; therefore, it should be administered into a rapidly flowing intravenous line. Dexrazoxane also can mitigate doxorubicin extravasation. Doxorubicin is metabolized by the liver, so doses must be reduced by 50–75% in the presence of liver dysfunction. Daunorubicin is closely related to doxorubicin and is preferable to doxorubicin owing to less mucositis and colonic damage with frequent high doses used in the

curative treatment of leukemia. Idarubicin is also used in leukemia treatment and may have somewhat less cardiotoxicity. Encapsulation of daunorubicin into a liposomal formulation has attenuated cardiac toxicity with antitumor activity in Kaposi's sarcoma, other sarcomas, multiple myeloma, and ovarian cancer.

Mitoxantrone is a synthetic topoisomerase II-directed agent with a mechanism similar to the anthracyclines, with less but not absent cardiotoxicity, comparing the ratio of cardiotoxic to effective doses; it is still associated with a 10% incidence of cardiotoxicity at cumulative doses of >150 mg/m². Etoposide binds directly to topoisomerase II and DNA in a reversible ternary complex. It stabilizes the covalent intermediate in the enzyme's action where the enzyme is covalently linked to DNA. Prominent clinical effects include myelosuppression, nausea, and transient hypotension related to the speed of administration of the agent. Etoposide is a mild vesicant but is relatively free from other large-organ toxicities.

Camptothecins target topoisomerase I. Topotecan is a camptothecin derivative approved for use in gynecologic tumors and small-cell lung cancer. Toxicity is limited to myelosuppression and mucositis. Irinotecan is a camptothecin with evidence of activity in colon carcinoma. Irinotecan is a prodrug, metabolized in the liver to SN-38, its active metabolite. Levels of SN-38 are particularly high in the setting of Gilbert's disease, characterized by defective uridine diphosphate glucuronosyl transferase (UGT) 1A1 and indirect hyperbilirubinemia, a

condition that affects about 10% of the white population in the United States. In addition, irinotecan's myelosuppression is clearly influenced by the patient's genotype for UGT1As. Irinotecan causes a delayed (48–72 h) secretory diarrhea related to the toxicity of SN-38. The diarrhea can be treated effectively with loperamide or octreotide; immediate diarrhea when it occurs is responsive to atropine.

Fam-trastuzumab deruxtecan and sacituzumab govitecan are antibody-drug conjugates (Fig. 73-4) that allow specific targeting of camptothecin and SN-38, respectively, to HER2-positive and triple-negative breast cancers, respectively. Adverse events are driven by off-target effects of the chemotherapy agent and include cytopenia, nausea, vomiting, and, in the case of fam-trastuzumab deruxtecan, severe interstitial pneumonitis.

Bleomycin forms complexes with Fe^{2+} while also bound to DNA. It remains an important component of curative regimens for Hodgkin's disease and germ cell neoplasms. Oxidation of Fe^{2+} gives rise to superoxide and hydroxyl radicals, causing DNA damage. The drug causes little, if any, myelosuppression. Bleomycin is cleared rapidly, but augmented skin and pulmonary toxicity in the presence of renal failure necessitates dose reduction in renal failure. Bleomycin is not a vesicant and can be administered intravenously, intramuscularly, or subcutaneously. Common side effects include fever and chills, facial flush, and Raynaud's phenomenon. The most feared complication of bleomycin treatment is pulmonary fibrosis, which increases in incidence at >300 cumulative units administered and is minimally responsive to treatment (e.g., glucocorticoids). The earliest indicator of an adverse effect is usually a decline in the carbon monoxide diffusing capacity (DLCO) or coughing, although cessation of drug immediately upon documentation of a decrease in DLCO may not prevent further decline in pulmonary function. Bleomycin is inactivated by a bleomycin hydrolase, whose concentration is diminished in skin and lung. Because bleomycin-dependent electron transport is dependent on O_2 , bleomycin toxicity may become apparent after exposure to transient very high fraction of inspired oxygen (FIO_2) even late after treatment. Thus, during surgical procedures, patients with prior exposure to bleomycin should be maintained on the lowest FIO_2 consistent with maintaining adequate tissue oxygenation.

Dactinomycin interacts directly with DNA to inhibit RNA transcription. It is important in the curative treatment of pediatric neoplasms, some of which also occur in young adults. Prominent myelosuppression, mucositis, alopecia, radiation recall, and nausea require management.

Calicheamicins are DNA-interacting antitumor antibiotics too toxic for clinical use but, when used as antibody-drug conjugates, can be useful in the treatment of CD33+ acute myeloid leukemia (gemtuzumab ozogamicin) and CD22+ acute lymphocytic leukemia (inotuzumab ozogamicin). Patients must be monitored for hypersensitivity reactions and for hepatotoxicity due to veno-occlusive disease of hepatic veins, resulting from release of the calicheamicin or metabolites in the liver.

Antimetabolites A broad definition of antimetabolites would include compounds that interfere with purine or pyrimidine synthesis. Some antimetabolites also cause DNA damage indirectly, through misincorporation into DNA. They tend to convey greatest toxicity to cells in S-phase, and the degree of toxicity increases with duration of exposure. Common toxic manifestations include stomatitis, diarrhea, and myelosuppression.

Methotrexate inhibits dihydrofolate reductase, which regenerates reduced folates from the oxidized folates produced when thymidine monophosphate is formed from deoxyuridine monophosphate. Without reduced folates, cells die a "thymine-less" death. N5-Tetrahydrofolate or N5-formyltetrahydrofolate (leucovorin) can bypass this block and rescue cells from methotrexate, which is retained in cells by polyglutamylation. Methotrexate is transported into cells by a membrane carrier, and high concentrations of drug can bypass this carrier and allow diffusion of drug directly into cells. These properties have suggested the design of "high-dose" methotrexate regimens with leucovorin rescue of normal marrow and mucosa as part of curative approaches to osteosarcoma in the adjuvant setting and hematopoietic

neoplasms of children and adults. Methotrexate is cleared by the kidney via both glomerular filtration and tubular secretion, and toxicity is augmented by renal dysfunction and drugs such as salicylates, probenecid, and nonsteroidal anti-inflammatory agents that undergo tubular secretion. With normal renal function, 15 mg/m² leucovorin will rescue 10^{-8} – 10^{-6} M methotrexate in 3–4 doses. However, with decreased creatinine clearance, doses of 50–100 mg/m² are continued until methotrexate levels are $<5 \times 10^{-8}$ M. In addition to bone marrow suppression and mucosal irritation, methotrexate can cause renal failure itself at high doses owing to crystallization in renal tubules; therefore, high-dose regimens require alkalinization of urine with increased flow by hydration. Methotrexate can be sequestered in third-space collections and diffuse back into the general circulation, causing prolonged myelosuppression. Less frequent adverse effects include reversible increases in transaminases and hypersensitivity-like pulmonary syndrome. Chronic low-dose methotrexate can cause hepatic fibrosis. When administered to the intrathecal space, methotrexate can cause chemical arachnoiditis and CNS dysfunction.

Pemetrexed is a folate-directed antimetabolite that inhibits the activity of several enzymes, including thymidylate synthetase (TS), dihydrofolate reductase, and glycinate ribonucleotide formyltransferase. To avoid toxicity to normal tissues, pemetrexed is given with low-dose folate and vitamin B₁₂ supplementation. Pemetrexed has notable activity against certain lung cancers and, in combination with cisplatin, also against mesotheliomas.

5-Fluorouracil (5-FU) represents an early example of "rational" drug design in that tumor cells incorporate radiolabeled uracil more efficiently into DNA than normal cells. 5-FU is metabolized in cells to 5'-FdUMP, which inhibits TS. In addition, misincorporation can lead to single-strand breaks, and RNA can aberrantly incorporate FUMP. 5-FU is metabolized by dihydropyrimidine dehydrogenase, and deficiency of this enzyme can lead to excessive toxicity from 5-FU. Oral bioavailability varies unreliable, but prodrugs such as capecitabine have been developed that allow at least equivalent activity to parenteral 5-FU-based approaches. Intravenous administration of 5-FU leads to bone marrow suppression after short infusions but to stomatitis after prolonged infusions. Leucovorin augments the activity of 5-FU by promoting formation of the ternary covalent complex of 5-FU, the reduced folate, and TS. Less frequent toxicities include CNS dysfunction, with prominent cerebellar signs, and endothelial toxicity manifested by thrombosis, including pulmonary embolus and myocardial infarction. Trifluridine is a fluorinated pyrimidine that as the triphosphate is directly incorporated into DNA, evoking DNA damage, and as the monophosphate can inhibit TS. It is administered as a fixed-dose combination with tipiracil, an inhibitor of trifluridine degradation by thymidine phosphorylase.

Cytosine arabinoside (ara-C) is incorporated into DNA after formation of ara-CTP, resulting in S-phase-related toxicity. Continuous infusion schedules allow maximal efficiency, with uptake maximal at 5–7 μM . Ara-C can be administered intrathecally. Adverse effects include nausea, diarrhea, stomatitis, chemical conjunctivitis, and cerebellar ataxia. Gemcitabine is a cytosine derivative that is similar to ara-C in that it is incorporated into DNA after anabolism to the triphosphate, rendering DNA susceptible to breakage and repair synthesis, which differs from that in ara-C in that gemcitabine-induced lesions are very inefficiently removed. In contrast to ara-C, gemcitabine appears to have useful activity in a variety of solid tumors, with limited nonmyelosuppressive toxicities.

6-Thioguanine and 6-mercaptopurine (6MP) are used in the treatment of acute lymphoid leukemia. Although administered orally, they display variable bioavailability. 6MP is metabolized by xanthine oxidase and therefore requires dose reduction when used with allopurinol. 6MP is also metabolized by thiopurine methyltransferase; genetic deficiency of thiopurine methyltransferase results in excessive toxicity.

Fludarabine phosphate is a prodrug of F-adenine arabinoside (F-ara-A), which in turn was designed to diminish the susceptibility of ara-A to adenosine deaminase. F-ara-A is incorporated into DNA and can cause delayed cytotoxicity even in cells with low growth fraction, including chronic lymphocytic leukemia and follicular B-cell

lymphoma. CNS and peripheral nerve dysfunction and T-cell depletion leading to opportunistic infections can occur in addition to myelosuppression. 2-Chlorodeoxyadenosine is a similar compound with activity in hairy cell leukemia. Hydroxyurea inhibits ribonucleotide reductase, resulting in S-phase block. It is orally bioavailable and useful for the acute management of myeloproliferative states.

Asparaginase is a bacterial enzyme that causes breakdown of extracellular asparagine required for protein synthesis in certain leukemic cells. This effectively stops tumor cell DNA synthesis, as DNA synthesis requires concurrent protein synthesis. The outcome of asparaginase action is therefore very similar to the result of the small-molecule antimetabolites. Because asparaginase is a foreign protein, hypersensitivity reactions are common, as are effects on organs such as pancreas and liver that normally require continuing protein synthesis. This may result in decreased insulin secretion with hyperglycemia, with or without hyperamylasemia and clotting function abnormalities. Close monitoring of clotting functions should accompany use of asparaginase. Paradoxically, owing to depletion of rapidly turning over anticoagulant factors, thromboses particularly affecting the CNS may also be seen with asparaginase.

Mitotic Spindle Inhibitors Microtubules form the mitotic spindle, and in interphase cells, they are responsible for the cellular “scaffolding” along which various motile and secretory processes occur. Microtubules are composed of repeating heterodimers of α and β isoforms of the protein tubulin. Vincristine binds to the tubulin heterodimer with the result that microtubules are disaggregated. This results in the block of growing cells in M-phase, where a structurally disordered mitotic spindle apparatus is a powerful proapoptotic signal (Fig. 73-3). Vincristine is metabolized by the liver, and dose adjustment in the presence of hepatic dysfunction is required. It is a powerful vesicant, and infiltration can be treated by local heat and infiltration of hyaluronidase. At clinically used intravenous doses, neurotoxicity in the form of glove-and-stockings neuropathy is frequent. Acute neuropathic effects include jaw pain, paralytic ileus, urinary retention, and the syndrome of inappropriate antidiuretic hormone secretion. Myelosuppression is not seen at conventional doses. Vinblastine is similar to vincristine, except that it tends to be more myelotoxic, with more frequent thrombocytopenia and also mucositis and stomatitis. Vinorelbine is a vinca alkaloid that appears to have differences in resistance patterns in comparison to vincristine and vinblastine; it may be administered orally.

The taxanes include paclitaxel and docetaxel. These agents differ from the vinca alkaloids in that the taxanes stabilize microtubules against depolymerization. The “stabilized” microtubules function abnormally and are not able to undergo the normal dynamic changes of microtubule structure and function necessary for cell cycle completion. Taxanes are among the most broadly active antineoplastic agents for use in solid tumors, with evidence of activity in ovarian cancer, breast cancer, Kaposi’s sarcoma, and lung tumors. They are administered intravenously, and their vehicles cause hypersensitivity reactions. Premedication with dexamethasone (8–16 mg orally or intravenously 12 and 6 h before treatment) and diphenhydramine (50 mg) and cimetidine (300 mg), both 30 min before treatment, decreases but does not eliminate the risk of hypersensitivity reactions to the paclitaxel vehicle. A protein-bound formulation of paclitaxel (called *nab-paclitaxel*) has at least equivalent antineoplastic activity and decreased risk of hypersensitivity reactions. Paclitaxel may also cause myelosuppression, neurotoxicity in the form of glove-and-stockings numbness, and paresthesia. Docetaxel causes comparable degrees of myelosuppression and neuropathy. Docetaxel uses a different vehicle that can cause fluid retention in addition to hypersensitivity reactions; dexamethasone premedication with or without antihistamines is also frequently used. Cabazitaxel is a taxane with somewhat better activity in prostate cancers than earlier generations of taxanes, perhaps due to superior delivery to sites of disease.

Epothilones represent a class of microtubule-stabilizing agents optimized for activity in taxane-resistant tumors. Ixabepilone has clear evidence of activity in breast cancers resistant to taxanes and

anthracyclines such as doxorubicin. Side effects include myelosuppression and peripheral sensory neuropathy. Eribulin is a microtubule-directed agent with activity in patients who have had progression of disease on taxanes. It alters dynamics of microtubule remodeling in cells.

Ado-trastuzumab emtansine is an antibody conjugate of the HER2/neu-directed trastuzumab and a highly toxic microtubule targeted drug (emtansine), which by itself is too toxic for human use; the antibody-drug conjugate shows valuable activity in patients with breast cancer who have developed resistance to the “naked” antibody. Brentuximab vedotin is an anti-CD30 antibody drug conjugate with the distinct microtubule poison dolastatin with activity in neoplasms such as Hodgkin’s lymphoma where the tumor cells frequently express CD30. Polatuzumab vedotin analogously targets CD79a in B-cell lymphomas. Enfortumab vedotin uses an antibody to NECTIN4 to target the vedotin “warhead” to urothelial neoplasms expressing that target. Belantamab mafodotin targets BCMA (B-cell maturation) expressed myeloma but using a distinct microtubule toxin, auristatin. Toxicity from these agents is driven by off-target effects of the microtubule agent and include myelosuppression and neuropathy, but belantamab mafodotin can cause ocular keratopathy, which requires monitoring.

CANCER MOLECULAR TARGETED THERAPY

Agents in this class share the characteristic that they are directed at specific cancer cell molecular targets important in the proliferation of tumors. While these agents can ultimately lead to tumor cell death, this occurs by altered regulation of a specific biochemical pathway affecting tumor cell susceptibility to apoptosis or growth arrest (Fig. 73-3).

Hormone Receptor-Directed Therapy Steroid hormone receptor-related molecules were arguably the first “molecular target” classes of anticancer drugs. When bound to their ligands, these receptors can alter gene transcription in hormone-responsive tissues. While in some cases, such as breast cancer, demonstration of the target hormone receptor is necessary for their use, in other cases such prostate cancer (androgen receptor) and lymphoid neoplasms (glucocorticoid receptor), the relevant receptor is always present in the tumor.

Glucocorticoids are generally given in “pulsed” high doses in leukemias and lymphomas, where they induce cell death in tumor cells. Cushing’s syndrome and inadvertent adrenal suppression on withdrawal from high-dose glucocorticoids can be significant complications, along with infections common in immunosuppressed patients, in particular *Pneumocystis* pneumonia, which classically appears a few days after completing a course of high-dose glucocorticoids.

Tamoxifen is a partial estrogen receptor antagonist; it antagonizes in breast tumors, mirroring its effect on breast tissue, but owing to agonistic activities in vascular and uterine tissue, side effects include increased risk of thromboembolic phenomena and a small increased incidence of endometrial carcinoma, which appears after chronic use (usually >5 years). Progestational agents—including medroxyprogesterone acetate, androgens including fluoxymesterone (Halotestin), and, paradoxically, estrogens—have approximately the same degree of activity in primary hormonal treatment of breast cancers that have elevated expression of estrogen receptor protein. Estrogen itself is not used often due to prominent cardiovascular and uterotrophic activity.

Aromatase refers to a family of enzymes that catalyze the formation of estrogen in various tissues, including ovary, peripheral adipose tissue, and some tumor cells. Aromatase inhibitors are of two types: irreversible steroid analogues such as exemestane and the reversible inhibitors such as anastrozole and letrozole. Anastrozole is superior to tamoxifen in the adjuvant treatment of breast cancer in postmenopausal patients with estrogen receptor-positive tumors. Letrozole treatment affords benefit following tamoxifen treatment. Adverse effects of aromatase inhibitors may include an increased risk of osteoporosis, fatigue, and altered serum lipids.

Metastatic prostate cancer is treated primarily by androgen deprivation. Orchiectomy causes responses in 80% of patients. If not accepted by the patient, testicular androgen suppression can also be induced by luteinizing hormone-releasing hormone (LHRH) agonists such as

leuprolide and goserelin. These agents cause tonic stimulation of the LHRH receptor, with loss of normal pulsatile activation resulting in net decreased output of luteinizing hormone (LH) by the anterior pituitary. Therefore, as primary hormonal manipulation in prostate cancer, one can choose orchiectomy or an LHRH agonist, but not both. This pathway can also be blocked by relugolix, an oral gonadotropin-releasing hormone antagonist.

The addition of androgen receptor blockers, including flutamide or bicalutamide, is of uncertain additional benefit in extending overall response duration, although pretreatment with these agents before LHRH agonists is important to avoid a surge in testosterone after initial LH release. Enzalutamide also binds to the androgen receptor and antagonizes androgen action in a mechanistically distinct way. Somewhat analogous to inhibitors of aromatase, agents have been derived that inhibit testosterone and other androgen synthesis in the testis, adrenal gland, and prostate tissue. Abiraterone inhibits 17 α -hydroxylase/C17,20 lyase (CYP17A1) and has been shown to be active in prostate cancer patients experiencing progression despite androgen blockade.

Tumors that respond to a primary hormonal manipulation may frequently respond to second and third hormonal manipulations. Thus, breast tumors that had previously responded to tamoxifen have, on relapse, notable response rates to withdrawal of tamoxifen itself or to subsequent addition of an aromatase inhibitor or progestin. Likewise, initial treatment of prostate cancers with leuprolide plus flutamide may be followed after disease progression by response to withdrawal of flutamide. These responses may result from the removal of antagonists from mutant steroid hormone receptors that have come to depend on the presence of the antagonist as a growth-promoting influence.

Non-Receptor-Linked Tyrosine Kinase Antagonists **Table 73-6** lists currently approved non-hormone receptor pathway-directed molecularly targeted chemotherapy agents, with features of their use of import to the generalist, particularly in recognizing potential drug-induced morbidities and interactions with other classes of drugs. The basis for discovery of drugs of this type was the prior knowledge of oncogene-directed pathways driving tumor growth (Fig. 73-3). In most cases, non-receptor tyrosine kinases ultimately activate signaling through the RAF/MEK/MAP kinase cascade, in common with the receptor-linked tyrosine kinases. Diagnostic demonstration of an active non-receptor tyrosine kinase may guide selection of an agent. A repeated preclinical and clinical observation in a variety of tumor types is that mutational activation of the tyrosine kinase target induces a state of “oncogene addiction” on the part of the tumor. This then is the basis for a “synthetic lethal” effect of the kinase inhibitor with respect to tumor viability.

In hematologic tumors, the prototypic agent of this type is imatinib, which targets the ATP binding site of the p210^{bcr-abl} protein tyrosine kinase that is formed as the result of the chromosome 9;22 translocation producing the Philadelphia chromosome in chronic myeloid leukemia (CML). It has lesser activity in the blast phase of CML, where the cells may have acquired additional mutations in p210^{bcr-abl} itself or other genetic lesions. Its side effects are relatively tolerable in most patients and include hepatic dysfunction, diarrhea, and fluid retention. Rarely, patients receiving imatinib have decreased cardiac function, which may persist after discontinuation of the drug. The quality of response to imatinib enters into the decision about when to refer patients with CML for consideration of stem cell transplant approaches. Nilotinib is a tyrosine protein kinase inhibitor with activity against p210^{bcr-abl} but with increased potency and perhaps better tolerance by certain patients. Dasatinib, another inhibitor of the p210^{bcr-abl} oncproteins, also has activity against certain mutant variants of p210^{bcr-abl} that are refractory to imatinib and arise during therapy or are present de novo. Dasatinib also has inhibitory action against kinases belonging to the src tyrosine protein kinase family; this activity may contribute to its effects. The T315I mutant of p210^{bcr-abl} is resistant to imatinib, nilotinib, bosutinib, and dasatinib; ponatinib has activity in patients with this T315Ip210^{bcr-abl}, but ponatinib has noteworthy associated thromboembolic toxicity. Use of this class of targeted agents is thus critically guided

not only by the presence of the p210^{bcr-abl} tyrosine kinase, but also by the presence of specific mutations in the ATP binding site.

Janus kinases (JAK) 1 and 2 are mutated in certain myeloproliferative states; cytopenias and infrequent arrhythmias infrequently complicate the use of ruxolitinib, the prototypic JAK inhibitor. Bruton's tyrosine kinase (BTK) is an intrinsic component of B-cell antigen receptor signaling and therefore is activated in many types of proliferating B cells. Inhibitors of BTK, including ibrutinib, acalabrutinib, and zanubrutinib, have noteworthy activity in certain lymphomas. Cytopenias and cardiac arrhythmias can occur, along with propensity to infection (indeed, the BTK was discovered as deficient in congenital hypogammaglobulinemia, presenting with repeated infections in childhood). Initial use of the BTK inhibitors requires consideration of prophylaxis against tumor lysis syndrome in case of a robust lympholytic effect of the agent.

Receptor-Linked Tyrosine Kinase Antagonists Mutated EGFR drives a significant fraction of non-small-cell lung cancers (NSCLCs). Erlotinib and gefitinib are the prototypic EGFR antagonists that, in early clinical trials, showed evidence of responses in a small fraction of patients with NSCLC. Subsequent studies by clinical oncologists in an effort to understand the basis of these excellent responses found that the probability of response to the agents was markedly increased in patients with an activating EGFR mutation, and current practice now routinely profiles patients with NSCLC for the presence of sensitizing mutations of EGFR. Side effects were generally acceptable, consisting mostly of acneiform rash (treated with glucocorticoid creams and clindamycin gel) and diarrhea. Patients with activating mutations who initially responded to gefitinib or erlotinib but who then had progression of the disease then acquired additional mutations in the enzyme, analogous to the mutational variants responsible for imatinib resistance in CML. Subsequent generations of EGFR antagonists have activity against more uncommon mutants (osimertinib) or a biochemically irreversible mechanism (dacomitinib).

Mutated anaplastic lymphoma kinase (ALK) and activated RET oncogene likewise drive distinct fractions of NSCLCs. Crizotinib, alectinib, and lorlatinib target ALK, but have prominent adverse cardiac, metabolic, and, in the case of lorlatinib, pulmonary events. Selpercatinib targets RET in NSCLCs (and thyroid cancers) but also with the chance of cardiac and liver toxicity.

Steel factor, a blood cell precursor-related bone marrow growth factor, uses the KIT receptor tyrosine kinase. KIT and variants of the platelet-derived growth factor receptor (PDGFR) are expressed in gastrointestinal stromal sarcoma (GIST). In addition to anti-p210^{bcr-abl} kinase activity, imatinib also inhibits mutants of KIT and PDGFR. Imatinib has found clinical utility in GIST, a tumor previously notable for its refractoriness to chemotherapeutic approaches. Imatinib's degree of activity varies with the specific mutational variant of KIT or PDGFR present in a particular patient's tumor.

HER2-driven breast cancers may be usefully treated with lapatinib; diarrhea and cardiac dysfunction can occur. Neratinib or tucatinib may also be useful in HER2-positive breast cancers after trastuzumab has ceased to be of value; diarrhea and liver toxicity also require monitoring and management.

Alteration of fibroblast growth factor (FGF) signaling can contribute to the growth of urothelial carcinomas and cholangiocarcinomas. Erdafitinib and pemigatinib, respectively, may be of utility with careful attention to ocular toxicity and hyperphosphatemia; the latter is an “on-target” toxicity of disrupting FGF receptor signaling in the kidney. Likewise, gilteritinib is active against the FMS-like tyrosine kinase-3 (FLT3) mutated in a fraction of poor-prognosis (treated by conventional chemotherapy) acute myeloid leukemias (AMLS). Cardiac, hepatic, gastrointestinal, and neurologic adverse events can occur, along with “differentiation” of the AML cells with cytokine elaboration and pulmonary side effects, requiring management with steroids and potentially hydroxyurea.

The neurotropic tyrosine kinase receptor (NRTK) undergoes translocation with fusion to a variety of different partners to produce a family of chimeric proteins in a small fraction of a variety of solid

TABLE 73-6 Molecularly Targeted Agents^a

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Non-Receptor Tyrosine Kinase Antagonists			
Acalabrutinib	Bruton's tyrosine kinase; mantle cell lymphoma after one prior treatment; CLL/SLL	Cytopenias, opportunistic infections, atrial fibrillation/flutter	CYP3A4, avoid proton pump inhibitors (PPIs); stagger administration with H ₂ blockers
Bosutinib	Bcr-Abl fusion protein (CML); wild-type and imatinib-resistant mutants	Myelosuppression, hepatic, QTc prolongation	CYP3A4; avoid PPIs; stagger administration with H ₂ blockers
Dasatinib	Bcr-Abl fusion protein (CML/ALL); wild-type and imatinib-resistant mutants	Myelosuppression (bleeding, infection); pulmonary hypertension, CHF, fluid retention; QTc prolongation; caution with hepatic impairment	CYP3A4; avoid PPIs; stagger administration with H ₂ blockers
Ibrutinib	Bruton's tyrosine kinase; CLL/SLL; mantle cell lymphoma after CD20-directed therapy; Waldenström's	Nausea, anemia, neutropenia, thrombocytopenia, fatigue, musculoskeletal pain, stomatitis, hypertension, cardiac arrhythmias, tumor lysis syndrome	CYP3A4
Imatinib	Bcr-Abl fusion protein (CML/ALL); c-kit mutants, PDGFR variants (GI stromal tumor [GIST]; eosinophilic syndromes)	Nausea, periorbital edema, rare CHF, QTc prolongation, hypothyroid	Myelosuppression not frequent in solid tumor indications; co-administration with CYP3A4 inducers/inhibitors may require dose adjustment; if need anticoagulation, no warfarin; heparinoids favored
Nilotinib	Bcr-Abl fusion protein (CML) and some imatinib-resistant variants	CHF, hepatic, QTc, electrolyte abnormalities, increased lipase, hypothyroidism	Interaction with CYP3A4-metabolized drugs; also CYP2C8, CYP2C9, CYP2D6, and CYP2B6; avoid food 2 h before and 1 h after a dose
Ponatinib	T315I mutation of Bcr-Abl fusion protein (CML)	Clotting, hepatic, CHF, pancreatitis, neuropathy, rash, arrhythmia, tumor lysis, reversible posterior leukoencephalopathy, wound healing altered	CYP3A4
Ruxolitinib	Janus kinase 1,2; intermediate- or high-risk myelofibrosis, including primary myelofibrosis, post-polycythemia vera myelofibrosis and post-essential thrombocythemia myelofibrosis	Thrombocytopenia, anemia, dizziness, headache	Adjust dose in renal and hepatic impairment, strong CYP3A4 inhibitors, or with fluconazole >200 mg doses except with GVHD
Zanubrutinib	Bruton's tyrosine kinase; mantle cell lymphoma after one prior therapy	Cytopenia, cardiac arrhythmia	Avoid with CYP3A4 interacting agents
Receptor-Linked Tyrosine Kinase Antagonists			
Afatinib	First-line treatment of NSCLC with nonresistant ATP site mutation of EGFR	Diarrhea; rash; ocular keratitis; interstitial lung disease; hepatic failure	Dose adjustment with Pgp inhibitors
Alectinib	Anaplastic lymphoma kinase (ALK)-positive metastatic NSCLC	Hepatotoxicity; interstitial lung disease; renal impairment; bradycardia	Myalgia and CPK elevations with muscle pain, tenderness, weakness
Avapritinib	GIST unresectable or metastatic with a PDGFRA exon 18 mutation, including PDGFRA D842V mutations	Edema, nausea, fatigue, CNS effects including altered cognition, sleep and mood disorders, hallucinations	Monitor for intracranial hemorrhage; avoid CYP3A4 inducer/inhibitors
Ceritinib	ALK-positive NSCLC: advanced or metastatic	GI adverse reactions, may require dose adjustment; hepatotoxicity, hyperglycemia, interstitial lung disease (permanently discontinue); QT interval prolongation (monitor with concomitant drugs known to prolong QT)	CYP3A, CYP2C9
Crizotinib	ALK-positive NSCLC: advanced or metastatic	Interstitial pneumonitis; hepatic; QTc prolongation; bradycardia; visual loss	Avoid CYP3A4 inducer/inhibitor
Dacomitinib	Advanced or metastatic NSCLC with epidermal growth factor receptor (EGFR) exon 19 deletion or exon 21 L858R via irreversible mechanism	Diarrhea, cutaneous: hold and/or dose reduce; interstitial lung disease (permanently discontinue)	Avoid with PPIs; use locally acting antacids or H ₂ receptor antagonist and administer at least 6 h before or 10 h after H ₂ receptor antagonist; CYP2D6
Erdafitinib	Target FGFR; advanced or metastatic urothelial cancer with an FGFR3 or FGFR2 alteration that has progressed beyond traditional platinum-based therapies	Stomatitis, fatigue, cutaneous changes, diarrhea; uncommon central serous retinopathy; retinal detachment; therefore, monitor with ophthalmologic exams during treatment; hyperphosphatemia a pharmacodynamic effect due to FGF23/Klotho signaling disruption	CYP2C9, CYP3A4 interactors; OCT2 substrates; separate dosing by at least 6 h before or after administration of Pgp substrates
Erlotinib	First-line treatment of NSCLC with ATP site mutation of EGFR; second-line treatment of wild-type EGFR NSCLC; pancreatic cancer with gemcitabine	Rash, diarrhea, renal failure, interstitial pneumonitis, liver	Administer at least 1 h before or 2 h after meals; CYP3A4; avoid with PPIs and space dosing with antacids; can alter warfarin effect; microangiopathic hemolytic anemia especially in pancreatic cancer, rare
Gefitinib	First-line treatment of NSCLC with ATP site mutation of EGFR	Rash, diarrhea, rare interstitial pneumonitis, ocular keratitis, GI perforation	CYP3A4; avoid with PPIs; monitor warfarin effect with gefitinib. In the United States, only with prior documented benefit in second-line treatment of NSCLC if not EGFR mutated

(Continued)

TABLE 73-6 Molecularly Targeted Agents^a (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Gilteritinib	Relapsed or refractory AML with an <i>FLT3</i> mutation	Hepatotoxicity, myalgia/arthritis, fatigue/malaise, mucositis, edema, rash, noninfectious diarrhea, dyspnea, nausea, cough, constipation, eye disorders, hypotension, vomiting, and renal impairment	Also inhibits AXL; unusual AML differentiation syndrome, requiring corticosteroids and consideration of hydroxyurea; posterior reversible encephalopathy syndrome possible (discontinue); prolonged QT interval: interrupt and/or reduce dose with a QTcF >500 ms (correct hypokalemia or hypomagnesemia prior to and during administration); pancreatitis: interrupt and/or reduce dosage; Pgp substrates; CYP3A
Lapatinib	Breast cancer: with capecitabine in HER2/neu advanced/metastatic after trastuzumab and chemotherapy; with aromatase inhibition if ER positive, HER2/neu positive	↓LVEF; liver; rash; nausea; diarrhea; palmar-plantar erythrodysesthesia	Interstitial lung disease and pneumonitis (discontinue if severe); QTc: monitor ECG and electrolytes, CYP3A4, CYP2C8, Pgp substrate interactions
Larotrectinib	Targets TRKA, TRKB, and TRC fusion proteins; indicated in any adult or pediatric solid tumor with a neurotrophic receptor tyrosine kinase (NTRK) gene fusion without a known acquired resistance mutation, with no satisfactory alternative treatments, or that has progressed following treatment	Neurotoxicity with potential cognitive impairment; hepatotoxicity, modify dose or withhold depending on severity	CYP3A4
Lorlatinib	NSCLC: ALK-positive metastatic NSCLC that has progressed on crizotinib and at least one other ALK inhibitor; or with progression on alectinib or ceritinib as the first ALK inhibitor therapy for metastatic disease	Hyperlipidemia: initiate or increase the dose of lipid-lowering agents and withhold and resume or dose reduce based on severity; AV block: withhold and resume or dose modify; CNS effects including seizures, hallucinations, altered cognitive function, altered mood, suicidal ideation, altered speech, mental status, and sleep	Targets ALK and also anti-ROS activity but FDA label limited to ALK indications; CYP3A4 (NB severe hepatotoxicity with strong CYP3A inducers; discontinue strong CYP3A inducers for 3 plasma half-lives prior to use); interstitial lung disease (ILD): immediately withhold and consider discontinuance with suspected ILD/pneumonitis
Neratinib	Breast cancer: with capecitabine in HER2/neu advanced metastatic disease after two prior HER2/neu agents; extended adjuvant treatment after early-stage adjuvant trastuzumab	Diarrhea; nausea; vomiting; abdominal pain; increased ALT/AST	Aggressive diarrhea prophylaxis with loperamide; avoid concomitant PPI antacids; separate from administration of other antacids; avoid CYP3A4 concomitant medications
Osimertinib	First-line treatment of metastatic NSCLC with <i>EGFR</i> exon 19 deletions or exon 21 L858R mutations; <i>EGFR</i> T790M mutation-positive NSCLC, progressed on or after EGFR TKI therapy	Interstitial lung disease, QTc prolongation, cardiomyopathy, ocular keratitis	Avoid or adjust dose with strong CYP3A4 inducers
Pemigatinib	Cholangiocarcinoma: previously treated, unresectable, locally advanced or metastatic cholangiocarcinoma with a fibroblast growth factor receptor 2 (FGFR2) fusion or other rearrangement	Hyperphosphatemia as pharmacodynamic effect: adjust dose if needed; stomatitis, nausea, diarrhea.	Retinal detachment: perform ocular exam with ocular coherence tomography prior to and every 2–3 months during treatment; CYP3A4
Selpercatinib	NSCLC: advanced or metastatic and <i>RET</i> fusion positive; medullary thyroid cancer: advanced or metastatic <i>RET</i> -mutant medullary thyroid cancer; advanced or metastatic <i>RET</i> -fusion-positive thyroid cancer that requires systemic therapy and that is radioactive iodine refractory (if appropriate)	Hepatotoxicity: monitor liver functions every 2 weeks during first 3 months, then monthly; hypertension, wound healing effects: withhold 1 week prior to surgery and at least 2 weeks after surgery; hemorrhage	QT interval prolongations: assess QTc at baseline, maintain electrolytes; avoid with QTc-prolonging drugs, avoid with antacids, but if not avoidable, take with food (with PPI) or modify administration time (with H ₂ receptor antagonist or locally acting antacid); CYP3A, CYP2C8 interaction
Tucatinib	Breast cancer: with trastuzumab and capecitabine after one or more HER2/neu regimens in the metastatic setting	Diarrhea, hepatotoxicity	CYP3A4, CYP2C8 interaction

RAF/MEK Inhibitors

Binimetinib	In combination with encorafenib, for the treatment of patients with unresectable or metastatic melanoma with a <i>BRAF</i> V600E or V600K mutation	Cardiomyopathy, venous thromboembolism; ocular, interstitial lung disease, hepatotoxicity, rhabdomyolysis	Targets MEK; dose modify with liver disease
Cobimetinib	In combination with vemurafenib for unresectable or metastatic melanoma with a <i>BRAF</i> V600E or V600K mutation	New primary malignancies, cutaneous and noncutaneous; hemorrhage, retinal vein occlusion, cardiomyopathy: evaluate LVEF before and during treatment, severe dermatologic reactions, rhabdomyolysis, hepatotoxicity, photosensitivity	CYP3A interaction

(Continued)

TABLE 73-6 Molecularly Targeted Agents^a (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Dabrafenib	<i>BRAF</i> V600E in melanoma; both alone and in combination with trametinib; may be useful in other tumors with <i>BRAF</i> V600E	As a single agent : hyperkeratosis, headache, pyrexia, arthralgia, papilloma, alopecia, and palmar-plantar erythrodysesthesia syndrome; in combination with trametinib: pyrexia, chills, fatigue, rash, nausea, vomiting, diarrhea, abdominal pain, peripheral edema, cough, headache, arthralgia, night sweats, decreased appetite, constipation, and myalgia	New primary cutaneous malignancies; hemorrhagic events as single agent; CYP3A4, CYP2C8, CYP2C19, and CYP2B6 interactions
Encorafenib	<i>BRAF</i> V600E in melanoma (in combination with binimetinib)	Uveitis, hemorrhage, QTc prolongation, fatigue, nausea, vomiting	CYP3A4 interactions; avoid with hormonal contraceptives
Trametinib	<i>BRAF</i> V600E in melanoma (both as single agent and in combination with dabrafenib)	Rash, diarrhea, lymphedema; cardiomyopathy, ocular toxicity including retinal vein occlusion, interstitial lung disease, fever, hemorrhage, venous thromboembolism, hyperglycemia	In combination with dabrafenib: second neoplasms, hemorrhage, venous thrombosis, CHF, ocular, hyperglycemia; avoid CYP3A4, CYP2C8, CYP2C9, CYP2C19, or CYP2B6 interacting drugs
Vemurafenib	<i>BRAF</i> V600E in melanoma; alone and in combination with cobimetinib; may be useful in other tumors with <i>BRAF</i> V600E	Cutaneous squamous cell carcinoma, severe rash including Stevens-Johnson, allergic hypersensitivity, QTc prolongation, hepatic, ocular, photosensitivity	Usually combined with cobimetinib in melanoma; CYP3A4, CYP1A2, and CYP2D6 interactions
Apoptosis Modulation			
Venetoclax	Targets BCL2; indicated in CLL/SLL; AML: in combination with azacitidine or decitabine or low-dose cytarabine in treatment of newly diagnosed AML in adults who are age 75 years or older or who have comorbidities that preclude use of intensive induction chemotherapy	Neutropenia; infection: withhold for grade 3 and higher	Tumor lysis syndrome (TLS): anticipate TLS, assess risk in all patients. Premedicate with anti-hyperuricemics and ensure adequate hydration, with more intensive measures (intravenous hydration, frequent monitoring, hospitalization) as overall risk increases. Immunization: No live attenuated vaccines prior to, during, or after venetoclax treatment; CYP3A, Pgp interaction; take Pgp substrates at least 6 h before venetoclax
Multikinase Inhibitors			
Axitinib	Renal cell carcinoma, second line	HBP, hemorrhage/clotting; diarrhea, other GI including GI perforation, fatigue, hand-foot syndrome, hypothyroidism, reversible posterior leukoencephalopathy, proteinuria	Targets VEGFR, PDGFR, KIT; CYP3A4/5 interaction
Brigatinib	Advanced or metastatic ALK-positive NSCLC progressed on or intolerant to crizotinib	Interstitial lung disease, bradycardia, hypertension, visual disturbances, hyperglycemia, creatine phosphokinase elevations	Targets ALK and EGFR; CYP3A interaction; hormonal contraceptives may be ineffective due to decreased exposure as CYP3A4 substrates
Cabozantinib	Medullary thyroid cancer; renal cell cancer; hepatocellular carcinoma after sorafenib	Hypertension, thrombotic events, diarrhea, fistula/GI perforation/wound healing, reversible posterior leukoencephalopathy, hemorrhage, palmar-plantar erythrodysesthesia	Targets VEGFR2, MET, AXL, RET; modify dose with CYP3A4 interactors
Capmatinib	NSCLC with MET exon 14 skipping	Interstitial lung disease, hepatic, photosensitivity	Targets MET; avoid with CYP3A4 interactors
Entrectinib	NSCLC: advanced and/or ROS1 positive; any solid tumors with an <i>NTRK</i> gene fusion without a known acquired resistance mutation, with metastasis, or in which surgical resection is likely to result in severe morbidity, or in tumors with progression following treatment or no satisfactory alternative therapy	CHF, CNS effect, skeletal fractures; hepatotoxicity: monitor liver tests, including ALT and AST, every 2 weeks during the first month of treatment, then monthly thereafter; withhold or permanently discontinue based on severity; hyperuricemia: assess serum uric acid levels prior to initiation and periodically during treatment	Targets <i>NTRK</i> gene fusion proteins; QT prolongation: assess with electrolytes at baseline and during treatment; vision disorders: withhold for new visual changes and consider ophthalmologic evaluation; CYP3A4 interaction: patients with BSA >1.50 m ² , reduce the dose of entrectinib if co-administration of moderate or strong CYP3A inhibitors and if BSA ≤1.50 m ² ; avoid entrectinib; avoid with moderate and strong CYP3A inducers
Fedratinib	Intermediate-2 or high-risk primary or secondary (post-polycythemia vera or post-essential thrombocythemia) myelofibrosis	Anemia, thrombocytopenia, nausea, vomiting, diarrhea, hepatic, amylase/lipase, encephalopathy: check thiamine levels prior, replete if deficient	Targets Janus kinase 2, and FLT3, RET; CYP3A4, CYP2C19 interaction
Lenvatinib	Iodine-refractory differentiated thyroid cancer; with everolimus for renal cell carcinoma after one prior antiangiogenic; hepatocellular carcinoma; with pembrolizumab, for the treatment of advanced endometrial carcinoma that is not MSI-H or dMMR and disease progression following prior systemic therapy; candidates for curative surgery	Hypertension, cardiac dysfunction, arterial thromboembolism, hepatic, renal, proteinuria, diarrhea, fistula/GI perforation/wound healing, QTc prolongation, hypocalcemia, reversible posterior leukoencephalopathy, hemorrhage, altered thyroid	Targets VEGFR1/2/3, FGFR1/2/3/4, PDGF α , KIT, and RET

(Continued)

TABLE 73-6 Molecularly Targeted Agents^a (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Midostaurin	Newly diagnosed <i>FLT3</i> -mutated AML during induction, consolidation, with daunorubicin/cytarabine-based chemotherapy; aggressive systemic mastocytosis, mast cell leukemia; systemic mastocytosis with associated hematologic malignancy	Interstitial lung disease; nausea; diarrhea	Targets mutant <i>FLT3</i> , protein kinase C, and many other protein kinases
Pazopanib	Renal cell carcinoma, soft tissue sarcoma (not GIST or adipocytic)	Fatigue, diarrhea/GI, hypertension; arterial and venous thrombosis with embolism, hemorrhage; hepatotoxicity: potentially severe/fatal; measure liver chemistries before and during treatment; GI perforation or fistula; proteinuria: monitor urine protein and interrupt treatment for 24-h urine protein ≥ 3 g and discontinue for repeat episodes despite dose reductions; infection: serious infections (with or without neutropenia); hypothyroidism	Targets VEGFRs, KIT, PDGFR, and FGFR; CHF \pm prolonged QT intervals and torsades des pointes: monitor LVEF, ECG, and electrolytes at baseline and during treatment; reversible posterior leukoencephalopathy syndrome, interstitial lung disease/pneumonitis, thrombotic microangiopathy, including thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome (permanently discontinue); CYP3A4, CYP2D6, CYP2C8 interaction; use with simvastatin increases the risk of ALT elevations and should be undertaken with caution; avoid with drugs that raise gastric pH; consider short-acting antacids in place of PPIs and H ₂ receptor antagonists; separate antacid and pazopanib dosing by several hours
Pexidartinib	Indicated for tenosynovial giant cell tumor (TGCT) associated with severe morbidity or functional limitations and not amenable to improvement with surgery	Administer 1 h before or 2 h after food; can cause serious and potentially fatal liver injury; monitor liver tests prior to and during treatment and withhold, dose reduce, or permanently discontinue	Targets colony-stimulating factor-1 receptor, KIT, <i>FLT3</i> ; avoid with agents known to cause hepatotoxicity; CYP3A, UGT interaction; avoid with PPIs; use H ₂ receptor antagonists or antacids if needed
Regorafenib	Second-line colorectal cancer; GI stromal tumor	Hypertension, hemorrhage, hand-foot syndrome and other dermatologic toxicity, thromboses, GI perforations with fistula, wound healing delays	Targets VEGFR, cardiac ischemia with infarction, reversible posterior leukoencephalopathy syndrome; CYP3A4 interaction
Sorafenib	Renal cell, hepatocellular, differentiated thyroid carcinoma	Diarrhea, hemorrhage, hand-foot syndrome, other rash, hypertension, CHF, QTc prolongation, hepatic toxicity, GI perforation	Targets c-RAF more selectively than B-RAF; VEGFR; many other kinases; impaired TSH suppression in thyroid cancer; CYP3A4 interaction
Sunitinib	Renal cell carcinoma, advanced or adjuvant; pancreatic neuroendocrine tumor, GIST after imatinib	Hypertension, hemorrhagic events, GI perforation, proteinuria, leading to renal failure: interrupt treatment for 24-h urine protein ≥ 3 g; discontinue for repeat episodes despite dose reductions; thyroid dysfunction, hypoglycemia: check blood glucose levels and consider antidiabetic drug dose modifications; osteonecrosis of the jaw: consider preventive dentistry prior to treatment and avoid invasive dental procedures, particularly in patients receiving intravenous bisphosphonate therapy; impaired wound healing: temporary interruption prior to major surgical procedures; palmar-plantar erythrodysthesia	Targets VEGFRs; PDGFR, RET, KIT; other protein kinases. Rare prolonged QT intervals and torsades des pointes: monitor at baseline and during treatment; maintain K, Mg levels; rare tumor lysis syndrome reported primarily in patients with RCC and GIST with high tumor burden; rare thrombotic microangiopathy, including thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome (discontinue); rare necrotizing fasciitis; severe cutaneous adverse events including erythema multiforme, Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN); discontinue if these events; CYP3A4 interaction
Vandetanib	Medullary thyroid cancer	Diarrhea, rash, hypertension, prolonged QTc, thromboses, fistulas, osteonecrosis, proteinuria	Targets VEGFR, RET, EGFR; CYP3A4 interaction

Cyclin-Dependent Kinase (CDK) Inhibitors

Abemaciclib	Breast cancer: with an aromatase inhibitor as initial endocrine-based therapy for the treatment of postmenopausal HR+, HER2- advanced or metastatic breast cancer; or with fulvestrant for the treatment of women with HR+, HER2- advanced or metastatic breast cancer with disease progression following endocrine therapy; as monotherapy for the treatment of adult patients with HR+, HER2- advanced or metastatic breast cancer with disease progression following endocrine therapy and prior chemotherapy	Diarrhea, neutropenia, thrombocytopenia, hepatotoxicity, venous thromboembolism	Targets CDK4/6; avoid concomitant use of ketoconazole; CYP3A4 interaction
-------------	--	---	---

(Continued)

TABLE 73-6 Molecularly Targeted Agents^a (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Palbociclib	Breast cancer: HR+, HER2– advanced or metastatic breast cancer in combination with an aromatase inhibitor as initial endocrine-based therapy in postmenopausal women; or fulvestrant in women with disease progression following endocrine therapy	Neutropenia, anemia, thrombocytopenia, stomatitis, diarrhea, fatigue	Targets CDK4/6; CYP3A interaction
Ribociclib	Breast cancer: with letrozole as initial endocrine-based therapy for the treatment of postmenopausal women with HR+, HER2– advanced or metastatic breast cancer	Hepatotoxicity; neutropenia	Targets CDK4/6; unusual QT interval prolongation; drugs known to prolong QT interval should be avoided such as antiarrhythmics; CYP3A interaction
Protein Homeostasis Modulators			
Bortezomib	Multiple myeloma, mantle cell lymphoma, second line	Neuropathy, thrombocytopenia, neutropenia, nausea, diarrhea, hypotension, tumor lysis syndrome with high tumor burden; hepatic: monitor hepatic enzymes during treatment, consider interruption	Proteasome inhibitor; infiltrative pulmonary disease, reversible posterior leukoencephalopathy syndrome: consider MRI for onset of visual or neurologic symptoms and discontinue if suspected; thrombotic microangiopathy; CYP3A4 interaction
Carfilzomib	Multiple myeloma: with dexamethasone or with lenalidomide plus dexamethasone in patients with relapsed or refractory multiple myeloma who have received one to three lines of therapy, or as a single agent for the treatment of patients with relapsed or refractory multiple myeloma who have received one or more lines of therapy	Infusion reaction: premedicate with dexamethasone; thrombocytopenia; tumor lysis syndrome, with need for hydration, monitoring of metabolic parameters	Proteasome inhibitor; cardiac toxicities: including failure or ischemia, withhold and evaluate; acute renal failure: monitor serum creatinine regularly; pulmonary toxicity, including pulmonary hypertension, acute respiratory distress/failure and diffuse infiltrative pulmonary disease: withhold and evaluate promptly; dose adjust with hepatic impairment; administer after a hemodialysis procedure
Ixazomib	Multiple myeloma: with lenalidomide and dexamethasone after at least one prior therapy	Thrombocytopenia, nausea, diarrhea, peripheral neuropathy, edema, hepatotoxicity	Proteasome inhibitor; avoid with strong CYP3A4 inducers; dose adjust with hepatic or renal impairment
Selinexor	Multiple myeloma (refractory): with dexamethasone after at least four prior therapies and refractory to at least two proteasome inhibitors, at least two immunomodulatory agents, and an anti-CD38 monoclonal antibody; DLBCL (relapsed or refractory) or arising from FL after at least two lines of systemic therapy	Thrombocytopenia, neutropenia, nausea, diarrhea, hyponatremia, neurotoxicity	Targets exportin 1 and therefore decreases efficient transport of proteins from nucleus to cytoplasm, leading to cell cycle arrest
Chromatin-Modifying Epigenetic Modulators			
DNA hypomethylating agents			
Azacitidine and decitabine	AML/myelodysplastic syndrome	Marrow, nausea, liver, neurologic, myalgia	“Suicide” inhibition of DNA methyl transferase after incorporation into DNA
Histone deacetylase inhibitors			
Belinostat	Peripheral T-cell lymphoma, relapsed or refractory	Thrombocytopenia, neutropenia, lymphopenia, anemia, infection, hepatotoxicity	Tumor lysis syndrome monitoring
Panobinostat	Multiple myeloma in combination with bortezomib and dexamethasone, in patients with multiple myeloma who have received at least two prior regimens, including bortezomib and an immunomodulatory agent.	Diarrhea, potentially severe, requiring prophylaxis; cardiac ischemic events, arrhythmias, hemorrhage, hepatotoxicity, cytopenias	CYP3A4, CYP2D6 interactions; avoid concomitant antiarrhythmic drugs/QT-prolonging drugs
Romidepsin	Cutaneous T-cell lymphoma, second line	QT prolongation, nausea, vomiting, cytopenias	Monitor QT at baseline and during treatment; monitor PT, INR with warfarin derivatives; CYP3A4 interaction
Vorinostat	Cutaneous T-cell lymphoma, second line	Fatigue, diarrhea, hyperglycemia, thrombocytopenia, embolism, GI bleeding, QT prolongation	Monitor QT at baseline and during treatment; monitor PT, INR with warfarin derivatives

(Continued)

TABLE 73-6 Molecularly Targeted Agents^a (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Histone methyltransferase inhibitors			
Tazemetostat	Epithelioid sarcoma: advanced or metastatic not eligible for surgical resection; FL with <i>EZH2</i> mutation after two prior therapies or any FL that has relapsed or is refractory to alternative therapies	Fatigue, nausea, constipation	Avoid CYP3A4 inducers/inhibitors; monitor for emergence of myelodysplastic syndrome, leukemia
Metabolism Modulators: mTOR Inhibitors/PI3 Kinase Inhibitors/IDH Inhibitors			
Alpelisib	Breast cancer: with fulvestrant for postmenopausal women, and men, with HR+, HER2-, <i>PIK3CA</i> -mutated, advanced or metastatic breast cancer following progression on or after an endocrine-based treatment	Hyperglycemia: safety not established in type 1 or uncontrolled type 2 diabetes; monitor glucose levels and hemoglobin A _{1c} ; optimize oral antihyperglycemics if warranted; interstitial pneumonitis: discontinue; diarrhea ≤ grade 2 frequent	Targets PI3K α isoform; severe hypersensitivity: permanently discontinue and initiate appropriate treatment; severe cutaneous reactions including SJS, erythema multiforme (EM), and TEN: consider consultation with a dermatologist; permanently discontinue if SJS, EM, or TEN confirmed; CYP3A4, CYP2C9; avoid with BCRP inhibitors
Copanlisib	Relapsed FL patients who have received at least two prior systemic therapies; pending confirmatory trial	Infection, hyperglycemia, HBP, noninfectious pneumonitis, neutropenia, cutaneous reactions	Targets PI3K α/δ isoforms; CYP3A4
Duvelisib	For relapsed or refractory CLL/SLL or FL; orphan drug designation for peripheral T-cell lymphoma	Neutropenia, hepatic toxicity, severe infections, diarrhea/colitis may require withholding; severe cutaneous reactions or pneumonitis in 5%	Targets PI3K γ/δ isoforms; CYP3A
Enasidenib	Relapsed or refractory AML with an <i>IDH2</i> mutation	Nausea, vomiting, diarrhea, elevated bilirubin, and anorexia	Targets <i>IDH2</i> mutant enzyme; unusual "differentiation syndrome" reflecting leukemia response to drug, but potentially fatal if not treated; use corticosteroid therapy, hemodynamic monitoring, consider hydroxyurea until symptom resolution
Everolimus	RCC, advanced; tuberous sclerosis-associated renal angiomyolipoma and/or subependymal giant cell astrocytoma; breast cancer, HR+, resistant to anastrozole or letrozole, in combination with exemestane; pancreatic, lung, or GI neuroendocrine, NOT functional carcinoid	Fatigue, noninfectious pneumonitis, infections, severe hypersensitivity reactions, renal impairment, impaired wound healing, hyperglycemia and hyperlipidemia, myelosuppression	Targets mTOR; angioedema with patients taking concomitant ACE inhibitors may be at increased risk; stomatitis: consider dexamethasone alcohol-free mouthwash when starting treatment; risk of reduced efficacy of vaccination: Pgp and strong CYP3A4 inhibitors: avoid concomitant use; Pgp and moderate CYP3A4 inhibitors: reduce dose; Pgp and strong CYP3A4 inducers: increase dose; geriatric patients: monitor and adjust dose for adverse reactions
Idelalisib	Relapsed CLL with rituximab; SLL, relapsed FL after two prior therapies	Hepatotoxicity, diarrhea or colitis, pneumonitis: monitor for pulmonary symptoms and bilateral interstitial infiltrates, then interrupt or discontinue; intestinal perforation: discontinue if suspected	Targets PI3K δ isoform; CYP3A4
Ivosidenib	AML: relapsed or refractory with an <i>IDH1</i> mutation; in the United States, newly diagnosed AML with a susceptible <i>IDH1</i> mutation, in patients who are at least 75 years old or who have comorbidities that preclude the use of intensive induction chemotherapy	Fatigue, leukocytosis, arthralgia, diarrhea, dyspnea, edema, nausea, mucositis, ECG QT prolonged, rash, pyrexia, cough, and constipation	Targets <i>IDH1</i> mutant; unusual QT prolongation (check electrolytes and hold or reduce dose); Guillain-Barré syndrome (permanently discontinue); CYP3A4; monitor/avoid with increased QTc-causing drugs
Temsirolimus	RCC, second line or poor prognosis	Hypersensitivity, hepatic (adjust dose in liver dysfunction), infection, interstitial lung disease, stomatitis, thrombocytopenia, nausea, anorexia, fatigue, hyperglycemia, hyperlipidemia, poor wound healing, GI perforation, renal impairment: check before treatment and periodically	Targets mTOR; CYP3A4/5 interactions; avoid live vaccines or exposure to subjects recently vaccinated with live vaccines
Poly-ADP Ribose Polymerase (PARP) Inhibitors			
Niraparib	Maintenance treatment of adult patients with recurrent epithelial ovarian, fallopian tube, or primary peritoneal cancer who are in a complete or partial response to platinum-based chemotherapy	Cytopenias, nausea, diarrhea, fatigue	Myelodysplastic syndrome

(Continued)

TABLE 73-6 Molecularly Targeted Agents^a (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Olaparib	Ovarian cancer: after two or more chemotherapies with deleterious <i>BRCA</i> mutation (germline and/or somatic); maintenance therapy when in complete or partial response to platinum-based chemotherapy Breast cancer: for the treatment of adult patients with deleterious or suspected deleterious g <i>BRCA</i> m, HER2– metastatic breast cancer who have been treated with chemotherapy in the neoadjuvant, adjuvant, or metastatic setting; if HR+, after prior endocrine therapy or inappropriate for endocrine therapy Pancreatic cancer: maintenance treatment of adult patients with deleterious or suspected deleterious g <i>BRCA</i> m metastatic pancreatic adenocarcinoma whose disease has not progressed on at least 16 weeks of a first-line platinum-based chemotherapy regimen	Nausea, fatigue, anemia, thrombocytopenia, neutropenia, stomatitis, liver function abnormalities	Myelodysplastic syndrome; rare interstitial pneumonitis
Rucaparib	Ovarian/fallopian tube/primary peritoneal cancer: as with olaparib	Nausea, fatigue, anemia, thrombocytopenia, neutropenia, stomatitis, liver function abnormalities	Severe heme toxicity with emergence of myelodysplastic syndrome
Talazoparib	Ovarian/fallopian tube/primary peritoneal cancer: as with olaparib	Nausea, fatigue (including asthenia), vomiting, abdominal pain, anemia, diarrhea, neutropenia, leukopenia, decreased appetite, constipation, stomatitis, dyspnea, and thrombocytopenia	Monitor for emergence of myelodysplasia; rare interstitial pneumonitis should lead to discontinuation; avoid with strong or moderate CYP3A inhibitors, but if concomitant use cannot be avoided, reduce dose; avoid with strong or moderate CYP3A inducers
Miscellaneous			
Arsenic trioxide	APL (target PML-RAR α and redox homeostasis)	\uparrow QT _c ; hypersensitivity with vasomotor symptoms	APL differentiation syndrome (see under tretinoin)
Glasdegib	AML: in combination with low-dose cytarabine, for the treatment of newly diagnosed AML in adult patients who are ≥ 75 years old or who have comorbidities that preclude use of intensive induction chemotherapy	Monitor ECG and electrolytes for QTc prolongation and interrupt treatment if it occurs	Targets smoothed receptor in hedgehog pathway; CYP3A4; avoid with QTc-prolonging drugs, but if co-administration is unavoidable monitor for increased QTc
Sonidegib	Metastatic basal cell carcinoma	Muscle spasm, fatigue, transmission through semen	Targets smoothed receptor in hedgehog pathway; CYP3A4
Tagraxofusp-erzs	Blastic plasmacytoid dendritic cell neoplasm	Hypersensitivity reactions (require premedication with steroids, antihistamines); hepatotoxicity, capillary leak syndrome	Targets CD123 (IL-3 receptor) to deliver a fragment of the diphtheria toxin
Tretinoin	APL, t(15;17) positive	Cutaneous including cheilitis, skin dryness; increased intracranial pressure; hyperlipidemia, abnormal liver function tests, usually resolve	Targets PML-RAR α ; APL differentiation syndrome: pulmonary dysfunction/infiltrate, pleural/pericardial effusion, fever
Vismodegib	Metastatic basal cell carcinoma	GI, hair loss, fatigue, muscle spasm, dysgeusia; no blood donation for 7 months after last dose	Targets smoothed receptor in hedgehog pathway
Ziv-aflibercept	Metastatic colorectal cancer in combination with 5-fluorouracil, leucovorin, irinotecan; resistant to or has progressed following an oxaliplatin-containing regimen	Fistula formation, GI perforation, hemorrhage, thrombosis, arterial thromboembolism, proteinuria, reversible posterior leukoencephalopathy	Targets VEGF by a solubilized receptor-trapping mechanism

^aAll agents in this category should be regarded as potentially fetotoxic, and use during pregnancy is either contraindicated or undertaken with clear understanding of risk of fetal harm; likewise not recommended for use during lactation.

Abbreviations: ACE, angiotensin-converting enzyme; ALL, acute lymphocytic leukemia; ALT, alanine aminotransferase; AML, acute myeloid leukemia; APL, acute promyelocytic leukemia; AST, aspartate aminotransferase; AV, atrioventricular; BCRP, breast cancer resistance protein drug transporter; BSA, body surface area; CHF, congestive heart failure; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; CNS, central nervous system; CPK, creatine phosphokinase; CYP, cytochrome p450 interactions with drugs metabolized by the indicated isoform; DLBCL, diffuse large B-cell lymphoma; dMMR, deficient mismatch repair; ECG, electrocardiogram; EGFR, epidermal growth factor receptor; ER, estrogen receptor; FDA, U.S. Food and Drug Administration; FL, follicular lymphoma; g*BRCA*m, germline mutated breast cancer associated protein; GI, gastrointestinal; GVHD, graft-versus-host disease; HBP, high blood pressure; MEK, mitogen activated protein kinase; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; IDH, isocitrate dehydrogenase; INR, international normalized ratio; LVEF, left ventricular ejection fraction; MSI-H, microsatellite instability-high; mTOR, mammalian target of rapamycin kinase; NSCLC, non-small-cell lung cancer; PDGFR, platelet-derived growth factor receptor; Pgp, P-glycoprotein; PT, prothrombin time; QTcF, QT corrected by Frederika formula; RCC, renal cell carcinoma; SLL, small lymphocytic lymphoma; TKI, tyrosine kinase inhibitor; TSH, thyroid-stimulating hormone; UGT, uridine diphosphate glucuronosyltransferase; VEGFR, vascular endothelial growth factor receptor.

tumors. Larotrectinib and entrectinib may be quite useful in managing these tumors; indeed, these agents are exemplary of “histology agnostic” agents, where the utility of the drug is not tied to a particular histologic diagnosis, but to the possession of a specific *NTRK* gene alteration. Neurotoxicity, a long half-life of the agents, and hepatotoxic adverse events are of concern. Also, assuring that solid tumors have been appropriately screened for the existence of such sensitizing mutations can be logistically and economically challenging.

RAS/RAF/MEK Antagonists The *BRAF* V600E mutation drives a substantial fraction of melanomas and certain NSCLCs and has been detected in certain thyroid tumors, colorectal tumors, hairy cell leukemias, and unusual gliomas. *BRAF* inhibitors such as dabrafenib, vemurafenib, and encorafenib have activity as single agents in many such tumors but are usually most active when co-administered as “doublets” with the MEK inhibitors trametinib, cobimetinib, and binimetinib, respectively, to promote “shut down” of RAF/MEK signaling at more than pathway member. Cutaneous adverse events including generally indolent cutaneous second neoplasms and thromboembolic, cardiac, and ocular toxicity can occur.

Sotorasib is a first-in-class inhibitor of *KRAS* G12C signaling that in early clinical reports has evidence of effecting stable disease in patients with a variety of neoplasm histologies bearing that mutation, with fewer actual responses. Its initial very favorable safety profile encourages further clinical investigations alone and in combination with other agents.

Multikinase Inhibitors Agents in this class also target specific macromolecules promoting the viability of tumor cells. They are “small-molecule” ATP site-directed antagonists that inhibit more than one protein kinase and may have value in the treatment of several solid tumors. Drugs of this type with prominent activity against the VEGFR tyrosine kinase have activity in renal cell carcinoma. Sorafenib is a VEGFR antagonist also with activity against the RAF serine-threonine protein kinase, and regorafenib is a closely related drug with value in relapsed advanced colon cancer. Pazopanib also prominently targets VEGFR and has activity in renal carcinoma and soft tissue sarcomas. Sunitinib has anti-VEGFR, anti-PDGFR, and anti-KIT activity. It causes prominent responses and stabilization of disease in renal cell cancers and GISTs. Side effects for agents with anti-VEGFR activity, similar to those of the anti-VEGF antibody bevacizumab, prominently include hypertension, proteinuria, and, more rarely, bleeding and clotting disorders, perforation of scarred gastrointestinal lesions, and posterior leukoencephalopathy, probably reflecting CNS vascular damage. Also encountered are fatigue, diarrhea, and hand-foot syndrome, with erythema and desquamation of the distal extremities, in some cases requiring dose modification, particularly with sorafenib.

Other agents in this class include agents such as brigatinib (clinical activity in ALK-dependent NSCLC, but also with anti-EGFR action), entrectinib (clinical activity in *NTRK* fusion protein diseases, but also in *ROS*-mutated NSCLC), and fedratinib (clinical activity in myeloproliferative neoplasms, but with RET activity in addition to JAK2 and FLT3 antagonism).

Cyclin-Dependent Kinase Inhibitors Cyclin-dependent kinases (CDKs) are activated as the result of oncogene pathway activity, and CDK4 and CDK6 phosphorylate the retinoblastoma (RB) tumor-suppressor gene to allow entry into S-phase. Palbociclib, abemaciclib, and ribociclib, selective inhibitors of CDK4 and CDK6, have noteworthy activity in advanced breast cancers also expressing the estrogen receptor, usually in conjunction with continued efforts to suppress estrogen receptor signaling, and frequently in conjunction with mTOR inhibitors. Further clinical investigations in other RB intact tumors may broaden their role.

Protein Homeostasis Modulators The proteasome is a macromolecular complex that degrades misfolded proteins tagged for removal by ubiquitin ligases. Proteasome inhibitors were originally designed as potential anti-inflammatory agents owing to proteasome activity to produce inflammatory cytokines but had unexpected

antiproliferative activity in a variety of cell types. Proteasome inhibitors have clinical utility in myeloma and lymphoma, where unbalanced synthesis of immunoglobulin components can accumulate after proteasome inhibitor treatment and induce apoptosis or starve cells for amino acids, inducing autophagy. Boronic acid proteasome inhibitors, including bortezomib and ixazomib, cause thrombocytopenia, gastrointestinal dysfunction, and neuropathy. Carfilzomib is a distinct chemotype with attenuated neuropathy but increased incidence of infusion reactions and cytokine release, with attendant risk of cardio-pulmonary adverse events.

Exportin 1 is a nuclear membrane transport protein that is responsible for normal exit and entry of a variety of nuclear proteins. Selinexor is an inhibitor of exportin action, resulting in abnormal nuclear accumulation of, e.g., tumor-suppressor gene products or needed export of other products, e.g., oncogene products. Useful clinical activity has been seen in myeloma and diffuse large B-cell lymphomas including those arising from previously treated indolent lymphomas. Cytopenias, gastrointestinal distress, and hyponatremia are features of its clinical use.

Chromatin-Modifying Agents Gene function is altered not only by mutation of DNA structure, but also by “epigenetic” mechanisms that alter the capacity of DNA to be transcribed or interact with regulatory proteins in the nucleus including transcription factors. Initial epigenetic approaches to modulate gene expression extended from the observation that low concentrations of certain nucleosides (5’azacytidine and decitabine) caused loss of methylated cytosine in DNA, associated with gene silencing, and had clinical activity in causing differentiation of AML cells with notably less toxicity than higher concentrations. 5’Azacytidine and decitabine are misincorporated into DNA and then scavenge DNA methyl transferase to disable DNA methylation of tumor-promoting genes and thus alter their transcription.

Histone deacetylase inhibitors alter the histone protein “packing” density of chromatin and induce global changes in expression of cell cycle regulatory proteins. Vorinostat, belinostat, and romidepsin are useful in cutaneous and peripheral T-cell lymphomas; panobinostat has activity in multiple myeloma. The agents are generally well tolerated but with the potential for cytopenias. The histone methyltransferase inhibitor tazemetostat is a first-in-class inhibitor of histone methyltransferase with unique activity in epithelioid sarcoma owing to its modulation of transcriptional mechanisms unique to that tumor and, recently, in certain follicular lymphomas.

Cancer Cell Metabolism Modulators Oncogenic transformation causes a “rewiring” of cellular metabolism away from oxidative phosphorylation to glycolysis (historically defined as the “Warburg effect” of aerobic glycolysis in animal and human tumors) with attendant tolerance of hypoxia and production of metabolites important for sustaining cell proliferation. Recent clinical studies have defined clinical value from inhibitors of the cell lipid membrane localized phosphoinositide-3 (PI3) kinase and mammalian target of rapamycin (mTOR) (the latter is a kinase whose inhibition was originally discovered as the mechanism by which the immunosuppressant rapamycin, isolated from a soil bacterium, decreased T-cell proliferation). PI3 kinase is activated by numerous oncogenic tyrosine kinases to ultimately cause a cascade of metabolic alterations including increased glucose uptake and activation of mTOR isoforms, which selectively increase translation efficiency of key regulators of cell cycle progression and protein synthetic capacity.

Temsirolimus and everolimus are mTOR inhibitors with activity in renal cancers. They produce stomatitis and fatigue; some hyperlipidemia (10%) and myelosuppression (10%); and rare lung toxicity and immunosuppression in regimens used clinically. Everolimus is also useful in patients with hormone receptor-positive breast cancers displaying resistance to hormonal inhibition and in certain neuroendocrine and brain tumors, the latter arising in patients with sporadic or inherited mutations in the pathway activating mTOR. Isoform-specific PI3 kinase inhibitors are of increasing importance in breast cancers with mutated *PI3K α* (alpelisib; hyperglycemia and cutaneous

eruptions can occur) or owing to selective use of PI3K δ by lymphoid tissues in lymphomas (idelalisib, copanlisib, and duvelisib).

Isocitrate dehydrogenase (IDH) inhibitors (ivosidenib specific for IDH1 and enasidenib specific for IDH2) have activity in tumors with IDH mutants (AML, cholangiocarcinomas) that generate the “oncometabolite” 2-hydroxyglutarate, which alters DNA and histone methyltransferase activity. The drugs thus function indirectly as epigenetic chromatin modulating agents through effects on cellular metabolism.

DNA Repair Pathway Modulators DNA repair systems act physiologically to lessen the impact of environmental genomic damaging agents and influence the susceptibility to certain chemotherapy agents. DNA repair enzyme mutations underlie inherited cancer susceptibility syndromes such as mutated *BRCA* tumor-suppressor gene-associated breast and ovarian cancers, among others.

Laboratory investigations revealed that poly-ADP ribose polymerase (PARP) acts as a synthetic lethal gene with mutations in the homologous recombination repair pathway, including the *BRCA* gene. PARP responds to detection of DNA lesions by creating chains of poly-ADP, which serve as scaffolds for the localization of DNA repair proteins still active even with mutated *BRCA* isoforms. However, without PARP activity, the scaffolds cannot form, and the DNA damage becomes lethal. This observation immediately suggested the potential utility of PARP inhibitors (e.g., olaparib) as treatments potentially useful for *BRCA*-induced tumors. Recently, PARP inhibitor utility has been extended to tumors that do not harbor *BRCA* mutations but have given evidence of responding to platinum drugs, as a way of extending the useful effect of the chemotherapy treatment. This finding underscores the likelihood that sensitivity to DNA-directed cytotoxic drugs on the part of a tumor is at least in part related to the drugs’ ability to take advantage of a sensitizing effect of a tumor’s endogenous DNA repair capacity.

Miscellaneous Targeted Therapies The t(15;17) chromosomal translocation is diagnostic of acute promyelocytic leukemia (APL), a subset of AML. The translocation produces a chimeric fusion protein joining the retinoic acid receptor (RAR) α to the transcription factor PML. The abnormal protein, encoding PML-RAR α , blocks differentiation of the cancer cells. All-trans-retinoic acid (ATRA) binds to the chimeric protein, releasing the block to differentiation inducing response in APL with fewer complications from cytopenias and disordered coagulation seen with cytotoxic agents. Its use can be attended by a “differentiation syndrome” characterized by cytokine release from and organ infiltration by the tumor cells. Pulmonary function can be severely compromised but is generally responsive to glucocorticoids. Increased intracranial pressure can occur from ATRA, and headache should occasion fundoscopic exam.

Arsenic trioxide was found empirically to also be of value in treating APL, and further study revealed that it also modulates PML-RAR α levels, along with decreasing the tolerance of APL cells for free radical damage, inducing apoptosis. The combination of arsenic trioxide and ATRA is productive of very high rates of complete remission in APL. Arsenic trioxide can cause lengthening of the QT interval, and careful attention to concomitant medications, Mg $^{2+}$, ionized Ca $^{2+}$, and K $^{+}$ is necessary during treatment.

The sonic hedgehog transcription factor pathway is regulated by the WNT ligands, which are active during embryonic and fetal life and in certain neoplasms. The sonic hedgehog inhibitors sonidegib and glassdegib are useful in non-surgically treatable cutaneous basal cell carcinomas and certain AMLs, respectively, where the pathway is active.

High-affinity binding to receptors on tumor cells can deliver toxins to tumor cells, exemplified by the IL-3-diphtheria toxin fusion protein tagraxofusp-erzs, targeting the IL-3 receptor (CD123) and useful in blastic plasmacytoid dendritic cell neoplasms. Capillary leak syndrome induced by the toxin component requires careful monitoring of fluid balance to avoid pulmonary dysfunction in particular. Specific receptors for cytokines and growth factors can also serve as “traps” to sequester needed growth factors. Ziv-aflibercept is not an antibody, but a solubilized VEGF receptor VEGF binding domain, and therefore

may have a distinct mechanism of action from bevacizumab, but with comparable side effects.

■ SYSTEMIC RADIATION THERAPY

Systemically administered isotopes of iodine have an important role in the treatment of thyroid neoplasms, owing to the selective upregulation of the iodide transporter in the tumor cell compartment. Likewise, isotopes of samarium and radium have been found useful in the palliation of bony metastases of prostate cancer owing to their selective deposition at the tumor-bone matrix interface. Antibody-radioisotope complexes such as Y⁹⁰-ibritumomab-tiuxetan target CD20, are useful in treating lymphoma, or an isotope can be complexed to a ligand for which the tumor has high affinity. The latter strategy is employed by Lu¹⁷⁷-dotatate, where an analog of somatostatin brings the lutetium isotope close to tumors such as somatostatin receptor-positive gastroenteropancreatic neuroendocrine tumors.

RESISTANCE TO CANCER TREATMENTS

Resistance mechanisms to the conventional cytotoxic agents were initially characterized in the late twentieth century as defects in drug uptake, metabolism, or export by tumor cells. The *multidrug resistance (MDR)* gene, encoding P-glycoprotein (Pgp), is prototypic of transport proteins that efficiently excrete many drugs from tumor cells; no clinically useful modulator of this process has yet emerged. Drug-metabolizing enzymes such as cytidine deaminase are upregulated in resistant tumor cells, and this is the basis for so-called “high-dose cytarabine” regimens in the treatment of leukemia. Another resistance mechanism defined during this era involved increased expression of a drug’s target, exemplified by amplification of the dihydrofolate reductase gene, in patients who had lost responsiveness to methotrexate, or mutation of topoisomerase II in tumors that relapsed after topoisomerase II modulator treatment.

A second class of resistance mechanisms involves loss of the cellular apoptotic mechanism activated after the engagement of a drug’s target by the drug. This occurs in a way that is heavily influenced by the biology of the particular tumor type. For example, decreased alkylguanine alkyltransferase expression defines a subset of glioblastoma patients with the prospect of enhanced benefit from treatment with temozolamide but has no value in predicting benefit from temozolamide in epithelial neoplasms. Likewise, ovarian cancers resistant to platinating agents have decreased expression of the proapoptotic gene *BAX*.

A related class of resistance mechanisms emerged from sequencing of the targets of agents directed at oncogenic kinases, revealing mutated targets, as described previously. This relates to the phenomenon of tumor heterogeneity. Tumors harbor distinct populations of subclones that arise during the process of carcinogenesis, sharing to variable degrees mutations that may promote the growth of some subclones, but that are absent or are no longer relevant to the growth of other subclones. Really useful targeted therapies address a target present in all subclones and to which all tumor subclones require for tumor growth.

Finally, other mechanisms of resistance to targeted agents include the upregulation of alternate means of activating the pathway targeted by the agent. Thus, melanomas initially responsive to *BRAF* V600E antagonists such as vemurafenib may reactivate RAF signaling by employing variant isoforms that can bypass the drug. Likewise, inhibition of HER2/neu signaling in breast cancer cells can lead to the emergence of variants with distinct ways of activating downstream effectors such as PI3 kinase.

SUPPORTIVE CARE DURING CANCER TREATMENT

■ MYELOSUPPRESSION

Cytotoxic chemotherapeutic agents almost invariably affect bone marrow function. Titration of this effect determines the tolerated dose of the agent on a given schedule. Polymorphonuclear leukocytes (PMNs; $t_{1/2} = 6\text{--}8$ h), platelets ($t_{1/2} = 5\text{--}7$ days), and red blood cells (RBCs; $t_{1/2} = 120$ days) have most, less, and least susceptibility, respectively, to usually administered cytotoxic agents. The nadir count of each cell type

in response to classes of agents is characteristic. Maximal neutropenia occurs 6–14 days after conventional doses of anthracyclines, anti-folates, and antimetabolites. Alkylating agents differ from each other in the timing of cytopenias. Nitrosoureas, DTIC, and procarbazine can display delayed marrow toxicity, first appearing 6 weeks after dosing.

Complications of myelosuppression result from the predictable sequelae of the missing cells' function. *Febrile neutropenia* refers to the clinical presentation of fever and <1500 granulocytes/ μ L. Management of febrile neutropenia is considered in *Chap. 74*. Transfusion of granulocytes has no role in the management of febrile neutropenia, owing to their exceedingly short half-life, mechanical fragility, and clinical syndromes of pulmonary compromise with leukostasis after their use. Instead, colony-stimulating factors (CSFs) are used to augment bone marrow production of PMNs. The American Society of Clinical Oncology has developed practice guidelines for the use of granulocyte CSF (G-CSF) and GM-CSF (*Table 73-7*).

TABLE 73-7 Indications for the Clinical Use of G-CSF or GM-CSF

Preventive Uses

With the first cycle of chemotherapy (so-called *primary CSF administration*)

Not needed on a routine basis

Use if the probability of febrile neutropenia is $\geq 20\%$

Use if patient has preexisting neutropenia or active infection

Age >65 years treated for lymphoma with curative intent or other tumors treated by similar regimens

Poor performance status

Extensive prior chemotherapy

Dose-dense regimens in a clinical trial or with strong evidence of benefit

With subsequent cycles if febrile neutropenia has previously occurred (so-called *secondary CSF administration*)

Not needed after short-duration neutropenia without fever

Use if patient had febrile neutropenia in previous cycle

Use if prolonged neutropenia (even without fever) delays therapy

Therapeutic Uses

Afebrile neutropenic patients

No evidence of benefit

Febrile neutropenic patients

No evidence of benefit

May feel compelled to use in the face of clinical deterioration from sepsis, pneumonia, or fungal infection, but benefit unclear

In bone marrow or peripheral blood stem cell transplantation

Use to mobilize stem cells from marrow

Use to hasten myeloid recovery

In acute myeloid leukemia

G-CSF of minor or no benefit

GM-CSF of no benefit and may be harmful

In myelodysplastic syndromes

Not routinely beneficial

Use intermittently in subset with neutropenia and recurrent infection

What Dose and Schedule Should Be Used?

G-CSF: 5 mg/kg per day subcutaneously

GM-CSF: 250 mg/m² per day subcutaneously

Pegfilgrastim: one dose of 6 mg 24 h after chemotherapy

When Should Therapy Begin and End?

When indicated, start 24–72 h after chemotherapy

Continue until absolute neutrophil count is 10,000/ μ L

Do not use concurrently with chemotherapy or radiation therapy

Abbreviations: CSF, colony-stimulating factor; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor.

Source: From the American Society of Clinical Oncology: J Clin Oncol 24:3187, 2006.

Dangerous degrees of thrombocytopenia do not frequently complicate the management of patients with solid tumors receiving cytotoxic chemotherapy (with the possible exception of certain carboplatin-containing regimens), but they are frequent in patients with certain hematologic neoplasms where marrow is infiltrated with tumor. Severe bleeding related to thrombocytopenia occurs with increased frequency at platelet counts <20,000/ μ L in patients with acute leukemia and <10,000/ μ L in patients with solid tumors and is prevalent at counts <5000/ μ L.

The precise "trigger" point at which to transfuse patients has been defined as a platelet count of 10,000/ μ L or less in patients without medical comorbidities that may increase the risk of bleeding. This issue is important not only because of the costs of frequent transfusion but also because unnecessary platelet transfusions expose the patient to the risks of alloimmunization and loss of value from subsequent transfusion, as well as the infectious and hypersensitivity risks inherent in any transfusion. Prophylactic transfusions to keep platelets >20,000/ μ L are reasonable in patients with leukemia who are stressed by fever or concomitant medical conditions. Careful review of medication lists to prevent exposure to nonsteroidal anti-inflammatory agents and maintenance of clotting factor levels adequate to support near-normal prothrombin and partial thromboplastin time tests are important in minimizing the risk of bleeding in the thrombocytopenic patient.

Anemia associated with chemotherapy can be managed by transfusion of packed RBCs. Transfusion is not undertaken until the hemoglobin falls to <80 g/L (8 g/dL), compromise of end-organ function occurs, or an underlying condition (e.g., coronary artery disease) calls for maintenance of hemoglobin >90 g/L (9 g/dL). Randomized trials in certain tumors have raised the possibility that erythropoietin (EPO) use may promote tumor cell survival.

■ NAUSEA AND VOMITING

The most common side effect of chemotherapy administration is nausea, with or without vomiting. Nausea may be acute (within 24 h of chemotherapy), delayed (>24 h), or anticipatory of the receipt of chemotherapy. Highly emetogenic drugs (risk of emesis >90%) include DTIC, cyclophosphamide at >1500 mg/m², and cisplatin; moderately emetogenic drugs (30–90% risk) include carboplatin, cytosine arabinoside (>1 g/m²), ifosfamide, conventional-dose cyclophosphamide, and anthracyclines; low-risk (10–30%) agents include 5-FU, taxanes, etoposide, and bortezomib, with minimal risk (<10%) afforded by treatment with antibodies, bleomycin, busulfan, fludarabine, and vinca alkaloids.

Serotonin antagonists (5-HT₃) and neurokinin 1 (NK1) receptor antagonists are useful in "high-risk" chemotherapy regimens. The combination acts at both peripheral gastrointestinal and CNS sites that control nausea and vomiting. For example, the 5-HT₃ blocker dolasetron, 100 mg intravenously or orally; dexamethasone, 12 mg; and the NK1 antagonist aprepitant, 125 mg orally, are combined on the day of administration of severely emetogenic regimens, with repetition of dexamethasone (8 mg) and aprepitant (80 mg) on days 2 and 3 for delayed nausea. Alternate 5-HT₃ antagonists include ondansetron, given as 0.15 mg/kg intravenously for three doses just before and at 4 and 8 h after chemotherapy; palonosetron at 0.25 mg over 30 s, 30 min before chemotherapy; and granisetron, given as a single dose of 0.01 mg/kg just before chemotherapy. Emesis from moderately emetic chemotherapy regimens may be prevented with a 5-HT₃ antagonist and dexamethasone alone for patients not receiving doxorubicin and cyclophosphamide combinations; the latter combination requires the 5-HT₃/dexamethasone/aprepitant on day 1, but aprepitant alone on days 2 and 3. Emesis from low-emetic-risk regimens may be prevented with 8 mg of dexamethasone alone or with non-5-HT₃, non-NK1 antagonist approaches including the following.

Antidopaminergic phenothiazines act directly at the chemoreceptor trigger zone (CTZ) in the brainstem medulla and include prochlorperazine (Compazine), 10 mg intramuscularly or intravenously, 10–25 mg orally, or 25 mg per rectum every 4–6 h for up to four doses;

and thiethylperazine, 10 mg by potentially all of the above routes every 6 h. Haloperidol is a butyrophenone dopamine antagonist given at 1 mg intramuscularly or orally every 8 h. Metoclopramide acts on peripheral dopamine receptors to augment gastric emptying and is used in high doses for highly emetogenic regimens (1–2 mg/kg intravenously 30 min before chemotherapy and every 2 h for up to three additional doses as needed); intravenous doses of 10–20 mg every 4–6 h as needed or 50 mg orally 4 h before and 8 and 12 h after chemotherapy are used for moderately emetogenic regimens. 5-9-Tetrahydrocannabinol (Marinol) is a rather weak antiemetic compared to other available agents, but it may be useful for persisting nausea and is used orally at 10 mg every 3–4 h as needed. Olanzapine, an “atypical antipsychotic” acting at multiple neurotransmitter receptors, may be of value, most clearly in cases refractory to the measures described above. Some practice guidelines have endorsed its earlier use in adults receiving highly emetogenic chemotherapy regimens in combination with an NK1 antagonist plus an HT3 antagonist plus dexamethasone.

■ DIARRHEA

Similar to the vomiting syndromes, chemotherapy-induced diarrhea may be immediate or can occur in a delayed fashion up to 48–72 h after the drugs. Careful attention to maintained hydration and electrolyte repletion, intravenously if necessary, along with antimotility treatments such as “high-dose” loperamide (4 mg at the first occurrence of diarrhea, with 2 mg repeated every 2 h until 12 h without loose stools, not to exceed a total daily dose of 16 mg), are appropriate. Octreotide (100–150 µg), a somatostatin analogue, or intralumenally acting opiate-based preparations may be considered for patients not responding to loperamide.

■ MUCOSITIS

Irritation and inflammation of the mucous membranes (mucositis) particularly afflicting the oral and anal mucosa, but potentially involving the entire gastrointestinal tract, may accompany cytotoxic chemotherapy. Topical therapies, including anesthetics and barrier-creating preparations, may provide symptomatic relief in mild cases. Palifermin, a keratinocyte growth factor and member of the fibroblast growth factor family, is effective in preventing severe mucositis in the setting of high-dose chemotherapy with stem cell transplantation for hematologic malignancies. It may also prevent or ameliorate mucositis from radiation.

■ ALOPECIA

Chemotherapeutic agents vary widely in causing alopecia, with anthracyclines, alkylating agents, and topoisomerase inhibitors reliably causing near-total alopecia when given at therapeutic doses. Antimetabolites are more variably associated with alopecia. Psychological support and the use of cosmetic resources are to be encouraged. “Chemo caps” that reduce scalp temperature to decrease the degree of alopecia are controversial during treatment with curative intent of neoplasms, such as leukemia or lymphoma, or in adjuvant breast cancer therapy. The richly vascularized scalp can certainly harbor micrometastatic or disseminated disease.

■ GONADAL DYSFUNCTION AND PREGNANCY

All cancer treatments described in this chapter should be regarded as potentially injurious to the developing fetus and to newborns via lactation. However, there are gradations to the degree of reproductive harm. All agents tend to have increased risk of adverse outcomes when administered during the first trimester, and strategies to delay chemotherapy, if possible, until after this milestone should be considered if the pregnancy is to continue to term. Patients in their second or third trimester can be treated with most regimens for the common

neoplasms afflicting women in their childbearing years, with the exception of antimetabolites, particularly antifolates, which have notable teratogenic or fetotoxic effects throughout pregnancy. The need for anticancer chemotherapy per se is infrequently a clear basis to recommend termination of a concurrent pregnancy, although each treatment strategy in this circumstance must be tailored to the individual needs of the patient.

Cessation of ovulation and azoospermia reliably result from regimens that contain alkylating agents and topoisomerase poison. The duration of these effects varies with age and sex. Sperm banking before treatment may be considered. Females experience amenorrhea with anovulation after alkylating agent therapy; egg preservation may be considered but may delay inception of urgent treatment. Recovery of normal menses is frequent if treatment is completed before age 30, but patients are unlikely to recover menses after age 35. Even those who regain menses usually experience premature menopause. Because the magnitude and extent of decreased fertility can be difficult to predict, patients should be counseled to maintain effective contraception, preferably by barrier means, during and after therapy. Resumption of efforts to conceive should be considered in the context of the patient’s likely prognosis. Hormone replacement therapy should be undertaken in women who do not have a hormonally responsive tumor. For patients who have had a hormone-sensitive tumor primarily treated by a local modality, conventional practice would counsel against hormone replacement, but this issue is under investigation.

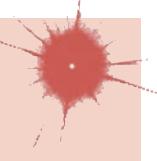
■ PALLIATIVE AND SUPPORTIVE CARE

An important perspective the primary care provider may bring to patients and their families facing incurable cancer is that, given the limited value of chemotherapeutic approaches at some point in the natural history of most metastatic cancers, *palliative care* or *hospice-based* approaches, with meticulous and ongoing attention to symptom relief and with family, psychological, and spiritual support, should receive prominent attention as a valuable therapeutic plan (Chaps. 12 and 69). Optimizing the quality of life rather than attempting to extend it becomes a valued intervention. Patients facing the impending progression of disease in a life-threatening way frequently choose to undertake toxic treatments of little to no potential value, and support provided by the primary caregiver in accessing palliative and hospice-based options in contrast to receiving toxic and ineffective regimens can be critical in providing a basis for patients to make sensible choices.

Late effects of cancer and its treatment are reviewed in Chap. 95.

■ FURTHER READING

- BROWN N et al: Precision medicine in non-small cell lung cancer: Current standards in pathology and biomarker interpretation. Am Soc Clin Oncol Educ Book 38:708, 2018.
- CHAN TA et al: Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. Ann Oncol 30:44, 2019.
- FORDE PM et al: Neoadjuvant PD-1 blockade in resectable lung cancer. N Engl J Med 378:1976, 2018.
- HESKETH PJ et al: Antiemetics: American Society of Clinical Oncology clinical practice update. J Clin Oncol 35:3240, 2017.
- NEELAPU SS et al: Chimeric antigen receptor T-cell therapy: Assessment and management of toxicities. Nat Rev Clin Oncol 15:47, 2018.
- PUZANOV I et al: Managing toxicities associated with immune checkpoint inhibitors: Consensus recommendations from the Society for Immunotherapy of Cancer (SITC) Toxicity Management Working Group. J Immunother Cancer 5:95, 2017.
- RIBAS A, WOLCHOK JD: Cancer immunotherapy using checkpoint blockade. Science 359:1350, 2018.
- YAP TA et al: The DNA damaging revolution: PARP inhibitors and beyond. Am Soc Clin Oncol Educ Book 39:185, 2019.



Infections are a common cause of death and an even more common cause of morbidity in patients with a wide variety of neoplasms. Infections in cancer patients can result directly from tissue invasion by cancerous cells (either by replacement of healthy host marrow cells or by occlusion of an orifice) (**Table 74-1**) or as a result of treatment. In the era of cytotoxic chemotherapy, neutropenia as a result of chemotherapy was the major cause of infectious complications of cancer therapy. The routine use of granulocyte-stimulating cytokines has, in most cases, shortened the duration of neutropenia, and the increasing use of checkpoint inhibitors and chimeric antigen receptor (CAR) T cells has changed the field of oncology and led to better outcomes. Unfortunately, checkpoint inhibitors and immunomodulators are also associated with an increased risk of infections—particularly intracellular pathogens. An evolving approach to prevention and treatment of infectious complications of cancer has decreased infection-associated mortality rates and will probably continue to do so. This accomplishment has resulted from three major steps:

- Early treatment:** The practice of using early empirical antibiotics reduced mortality rates among patients with leukemia and bacteremia from 84% in 1965 to 44% in 1972. The mortality rate due to infection in febrile neutropenic patients dropped to <10% by 2013. This dramatic improvement is attributed to early intervention with appropriate antimicrobial therapy.
- Empirical treatment:** “Empirical” antifungal therapy has also lowered the incidence of disseminated fungal infection, with dramatic decreases in mortality rates. An antifungal agent is administered—on the basis of likely fungal infection—to neutropenic patients who, after 4–7 days of antibiotic therapy, remain febrile but have no positive cultures.
- Prophylaxis:** Use of antibiotics for afebrile neutropenic patients as broad-spectrum prophylaxis against infections has decreased both mortality and morbidity even further. The current approach to treatment of severely neutropenic patients (e.g., those receiving high-dose chemotherapy for leukemia or high-grade lymphoma) is based on initial prophylactic therapy at the onset of neutropenia, subsequent “empirical” antibacterial therapy targeting the

organisms whose involvement is likely in light of physical findings (most often fever alone), and finally “empirical” antifungal therapy based on the known likelihood that fungal infection will become a serious issue after 4–7 days of broad-spectrum antibacterial therapy.

A physical predisposition to infection in patients with cancer (**Table 74-1**) can be a result of the neoplasm’s production of a break in the skin. For example, a squamous cell carcinoma may cause local invasion of the epidermis, which allows bacteria to gain access to subcutaneous tissue and permits the development of cellulitis. The artificial closing of a normally patent orifice can also predispose to infection; for example, obstruction of a ureter by a tumor can cause urinary tract infection, and obstruction of the bile duct can cause cholangitis. Part of the host’s normal defense against infection depends on the continuous emptying of a viscus; without emptying, a few bacteria that are present as a result of bacteremia or local transit can multiply and cause disease.

A similar problem can affect patients whose lymph node integrity has been disrupted by radical surgery, particularly patients who have had radical node dissections. A common clinical problem following radical mastectomy is the development of cellulitis (usually caused by streptococci or staphylococci) because of lymphedema and/or inadequate lymph drainage. In most cases, this problem can be addressed by local measures designed to prevent fluid accumulation and breaks in the skin, but antibiotic prophylaxis has been used in refractory cases.

A life-threatening problem common to many cancer patients is the loss of the reticuloendothelial capacity to clear microorganisms after splenectomy, which may be performed as part of the management of hairy cell leukemia, chronic lymphocytic leukemia (CLL), and chronic myelogenous leukemia (CML) as well as Hodgkin’s disease. Even after curative therapy for the underlying disease, the lack of a spleen predisposes such patients to rapidly fatal infections. The loss of the spleen through trauma similarly predisposes the normal host to overwhelming infection throughout life. The splenectomized patient should be counseled about the risks of infection with certain organisms, such as the protozoan *Babesia* (**Chap. 225**) and *Capnocytophaga canimorsus*, a bacterium carried in the mouths of animals (**Chaps. 141 and 158**). Because encapsulated bacteria (*Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis*) are the organisms most commonly associated with postsplenectomy sepsis, splenectomized persons should be vaccinated (and revaccinated; **Table 74-2** and **Chap. 123**) against the capsular polysaccharides of these organisms. Many clinicians recommend giving splenectomized patients a small supply of antibiotics effective against *S. pneumoniae*, *N. meningitidis*, and *H. influenzae* to avert rapid, overwhelming sepsis in the event that they cannot present for medical attention immediately after the onset of fever or other signs or

TABLE 74-1 Disruption of Normal Barriers in Patients with Cancer That May Predispose Them to Infections

TYPE OF DEFENSE	SPECIFIC LESION OR DEFICIENCY	CELLS INVOLVED	ORGANISM	CANCER ASSOCIATION	DISEASE
Physical barrier	Breaks in skin	Skin epithelial cells	Staphylococci, streptococci	Head and neck, squamous cell carcinoma	Cellulitis, extensive skin infection
Emptying of fluid collections	Occlusion of orifices: ureters, bile duct, colon	Luminal epithelial cells	Gram-negative bacilli	Renal, ovarian, biliary tree, metastatic diseases of many cancers	Rapid, overwhelming bacteremia; urinary tract infection
Lymphatic function	Node dissection	Lymph nodes	Staphylococci, streptococci	Breast cancer surgery	Cellulitis
Splenic clearance of microorganisms	Splenectomy	Splenic reticuloendothelial cells	<i>Streptococcus pneumoniae</i> , <i>Haemophilus influenzae</i> , <i>Neisseria meningitidis</i> , <i>Babesia</i> , <i>Capnocytophaga canimorsus</i>	Hodgkin’s disease, leukemia	Rapid, overwhelming sepsis
Phagocytosis	Lack of granulocytes	Granulocytes (neutrophils)	Staphylococci, streptococci, enteric organisms, fungi	Acute myeloid and acute lymphocytic leukemias, hairy cell leukemia	Bacteremia
Humoral immunity	Lack of antibodies	B cells	<i>S. pneumoniae</i> , <i>H. influenzae</i> , <i>N. meningitidis</i>	Chronic lymphocytic leukemia, multiple myeloma	Infections with encapsulated organisms, sinusitis, pneumonia
Cellular immunity	Lack of T cells	T cells and macrophages	<i>Mycobacterium tuberculosis</i> , <i>Listeria</i> , herpesviruses, fungi, intracellular parasites	Hodgkin’s disease, leukemia, T-cell lymphoma	Infections with intracellular bacteria, fungi, parasites; virus reactivation

*Deceased.

TABLE 74-2 Vaccination of Cancer Patients Receiving Chemotherapy^a

VACCINE	USE IN INDICATED PATIENTS		
	INTENSIVE CHEMOTHERAPY	HODGKIN'S DISEASE	HEMATOPOIETIC STEM CELL TRANSPLANTATION
Diphtheria-tetanus-pertussis ^b	Primary series and boosters as necessary	No special recommendation	3 doses given 6–12 months after transplantation
Poliomyelitis ^c	Complete primary series and boosters	No special recommendation	3 doses given 6–12 months after transplantation
<i>Haemophilus influenzae</i> type b conjugate	Primary series and booster for children	Single dose for adults	3 doses given 6–12 months after transplantation (separated by 1 month)
Human papillomavirus (HPV)	HPV vaccine is approved for males and females 9–26 years of age. Check Centers for Disease Control and Prevention (CDC) website (www.cdc.gov/vaccines) for updated recommendations.	HPV vaccine is approved for males and females 9–26 years of age. Check CDC website (www.cdc.gov/vaccines) for updated recommendations.	HPV vaccine is approved for males and females 9–26 years of age. Check CDC website (www.cdc.gov/vaccines) for updated recommendations.
Hepatitis A	As indicated for normal hosts on the basis of occupation and lifestyle	As indicated for normal hosts on the basis of occupation and lifestyle	As indicated for normal hosts on the basis of occupation and lifestyle
Hepatitis B	Same as for normal hosts	As indicated for normal hosts on the basis of occupation and lifestyle	3 doses given 6–12 months after transplantation
Pneumococcal conjugate vaccine (PCV13) Pneumococcal polysaccharide vaccine (PPSV23) ^d	Finish series prior to chemotherapy if possible.	Patients with splenectomy should receive both PCV13 and PPSV23.	Three doses of PCV13, beginning 3–6 months after transplantation, are followed by a dose of PPSV23 at least 8 weeks later. A second PPSV23 dose can be given 5 years later.
Quadrivalent meningococcal vaccine ^e	Should be administered to splenectomized patients and to patients living in endemic areas, including college students in dormitories	Should be administered to splenectomized patients and to patients living in endemic areas, including college students in dormitories. An additional dose can be given after 5 years.	Should be administered to splenectomized patients and to patients living in endemic areas, including college students in dormitories. An additional dose can be given after 5 years.
Meningococcal B vaccine	See above.	See above.	See above (see www.cdc.gov/vaccines for updated recommendations).
Influenza	Seasonal immunization	Seasonal immunization	Seasonal immunization (A seasonal dose is recommended and can be given as early as 4 months after transplantation; if given <6 months after transplantation, an additional dose is recommended.)
Measles/mumps/rubella	Contraindicated	Contraindicated during chemotherapy	After 24 months in patients without graft-versus-host disease
Varicella-zoster virus ^f	Zoster recombinant vaccine	Zoster recombinant vaccine	Two-dose zoster recombinant vaccine recommended

^aThe latest recommendations by the Advisory Committee on Immunization Practices and the CDC guidelines can be found at www.cdc.gov/vaccines. ^bA single dose of TDaP (tetanus–diphtheria–acellular pertussis), followed by a booster dose of Td (tetanus–diphtheria) every 10 years, is recommended for adults. ^cLive-virus vaccine is contraindicated; inactivated vaccine should be used. ^dTwo types of vaccines are used to prevent pneumococcal disease. A conjugate vaccine active against 13 serotypes (13-valent pneumococcal conjugate vaccine, or PCV13) is currently administered in three separate doses to all children. A polysaccharide vaccine active against 23 serotypes (23-valent pneumococcal polysaccharide vaccine, or PPSV23) elicits titers of antibody lower than those achieved with the conjugate vaccine, and immunity may wane more rapidly. Because the ablative chemotherapy given to recipients of hematopoietic stem cell transplants (HSCTs) eradicates immunologic memory, revaccination is recommended for all such patients. Vaccination is much more effective once immunologic reconstitution has occurred; however, because of the need to prevent serious disease, pneumococcal vaccine should be administered 6–12 months after transplantation in most cases. Because PPSV23 includes serotypes not present in PCV13, HSCT recipients should receive a dose of PPSV23 at least 8 weeks after the last dose of PCV13. Although antibody titers from PPSV23 clearly decay, experience with multiple doses of PPSV23 is limited, as are data on the safety, toxicity, or efficacy of such a regimen. For this reason, the CDC currently recommends the administration of one additional dose of PPSV23 at least 5 years after the last dose to immunocompromised patients, including transplant recipients, as well as patients with Hodgkin's disease, multiple myeloma, lymphoma, or generalized malignancies. Beyond this single additional dose, further doses are not recommended at this time. ^eMeningococcal conjugate vaccine (MenACWY) is recommended for adults ≤55 years old, and meningococcal polysaccharide vaccine (MPSV4) is recommended for those ≥56 years old. ^fVaricella vaccine is recommended for children and zoster recombinant vaccine for adults. ^gContact the manufacturer for more information on use in children with acute lymphocytic leukemia.

symptoms of bacterial infection. A few tablets of amoxicillin/clavulanic acid (or levofloxacin if resistant strains of *S. pneumoniae* are prevalent locally) are a reasonable choice for this purpose.

The level of suspicion of infections with certain organisms depends on the type of cancer diagnosed (Table 74-3). Diagnosis of multiple myeloma or CLL should alert the clinician to the possibility of hypogammaglobulinemia. While immunoglobulin replacement therapy can be effective, in most cases, prophylactic antibiotics are a cheaper, more convenient method of eliminating bacterial infections in CLL patients with hypogammaglobulinemia. Patients with acute lymphocytic leukemia (ALL), patients with non-Hodgkin's lymphoma, and all cancer patients treated with high-dose glucocorticoids (or glucocorticoid-containing chemotherapy regimens) should receive antibiotic prophylaxis for *Pneumocystis* infection (Table 74-3) for the duration of their chemotherapy. In addition to exhibiting susceptibility to certain infectious organisms, patients with cancer are likely to manifest their

infections in characteristic ways. For example, fever—generally a sign of infection in normal hosts—continues to be a reliable indicator in neutropenic patients. In contrast, patients receiving glucocorticoids and agents that impair T-cell function and cytokine secretion may have serious infections in the absence of fever. Similarly, neutropenic patients commonly present with cellulitis without purulence and with pneumonia without sputum or even x-ray findings (see below).

The use of monoclonal antibodies that target B and T cells as well as drugs that interfere with lymphocyte signal transduction events are associated with reactivation of latent infections. The use of infliximab and other anti-tumor necrosis factor (TNF) antibodies are associated with the development of reactivation tuberculosis. Similarly, the use of the anti-B cell antibody, rituximab, is associated with reactivation of hepatitis B and other latent viruses. Checkpoint inhibitors also predispose individuals to reactivation of intracellular pathogens, and clinicians must be aware of what viruses and other intracellular organisms

TABLE 74-3 Infections Associated with Specific Types of Cancer

CANCER	UNDERLYING IMMUNE ABNORMALITY	ORGANISM(S) CAUSING INFECTION
Multiple myeloma	Hypogammaglobulinemia	<i>Streptococcus pneumoniae</i> , <i>Haemophilus influenzae</i> , <i>Neisseria meningitidis</i>
Chronic lymphocytic leukemia	Hypogammaglobulinemia	<i>S. pneumoniae</i> , <i>H. influenzae</i> , <i>N. meningitidis</i>
Acute myeloid or lymphocytic leukemia	Granulocytopenia, skin and mucous membrane lesions	Extracellular gram-positive and gram-negative bacteria, fungi
Hodgkin's disease	Abnormal T-cell function	Intracellular pathogens (<i>Mycobacterium tuberculosis</i> , <i>Listeria</i> , <i>Salmonella</i> , <i>Cryptococcus</i> , <i>Mycobacterium avium</i>); herpesviruses
Non-Hodgkin's lymphoma and acute lymphocytic leukemia	Glucocorticoid chemotherapy, T- and B-cell dysfunction	<i>Pneumocystis</i>
Colon and rectal tumors	Local abnormalities ^a	<i>Streptococcus bovis</i> biotype 1 (bacteremia)
Hairy cell leukemia	Abnormal T-cell function	Intracellular pathogens (<i>M. tuberculosis</i> , <i>Listeria</i> , <i>Cryptococcus</i> , <i>M. avium</i>)

^aThe reason for this association is not well defined.

(mycobacteria, fungi, etc.) are likely to grow and pose a threat to an individual patient receiving these therapies. Like organ transplant recipients (Chap. 143), patients with latent bacterial disease (like tuberculosis) and latent viral disease (like herpes simplex or zoster) should be carefully monitored for reactivation disease.

SYSTEM-SPECIFIC SYNDROMES

SKIN-SPECIFIC SYNDROMES

Skin lesions are common in cancer patients, and the appearance of these lesions may permit the diagnosis of systemic bacterial or fungal infection. While cellulitis caused by skin organisms such as *Streptococcus* or *Staphylococcus* is common, neutropenic patients—that is, those with <500 functional polymorphonuclear leukocytes (PMNs)/μL—and patients with impaired blood or lymphatic drainage may develop infections with unusual organisms. Innocent-looking macules or papules may be the first sign of bacterial or fungal sepsis in immunocompromised patients (Fig. 74-1). In the neutropenic host, a macule progresses rapidly to ecthyma gangrenosum (see Fig. A1-34), a usually painless, round, necrotic lesion consisting of a central black or gray-black eschar with surrounding erythema. Ecthyma gangrenosum, which is located in nonpressure areas (as distinguished from necrotic lesions associated with lack of circulation), is often associated with *Pseudomonas aeruginosa* bacteremia (Chap. 164) but may be caused by other bacteria.

Candidemia (Chap. 216) is also associated with a variety of skin conditions (see Fig. A1-37) and commonly presents as a maculopapular rash. Punch biopsy of the skin may be the best method for diagnosis.

Cellulitis, an acute spreading inflammation of the skin, is most often caused by infection with group A *Streptococcus* or *Staphylococcus aureus*, virulent organisms normally found on the skin (Chap. 129). Although cellulitis tends to be circumscribed in normal hosts, it may spread rapidly in neutropenic patients. A tiny break in the skin may lead to spreading cellulitis, which is characterized by pain and erythema; in the affected patients, signs of infection (e.g., purulence) are often lacking. What might be a furuncle in a normal host may require amputation because of uncontrolled infection in a patient presenting with leukemia. A dramatic response to an infection that might be trivial in a normal host can mark the first sign of leukemia. Fortunately, granulocytopenic patients are likely to be infected with certain types of organisms (Table 74-4); thus, the selection of an antibiotic regimen is somewhat easier than it might otherwise be (see “Antibacterial Therapy” below). It is essential to recognize cellulitis early and to treat it aggressively. Patients who are neutropenic or who have previously



A



B

FIGURE 74-1 **A.** Papules related to *Escherichia coli* bacteremia in a patient with acute lymphocytic leukemia. **B.** The same lesions on the following day.

TABLE 74-4 Organisms Likely to Cause Infections in Granulocytopenic Patients

Gram-Positive Cocci	
<i>Staphylococcus epidermidis</i> ^a	<i>Staphylococcus aureus</i>
<i>Viridans Streptococcus</i>	<i>Enterococcus faecalis</i>
<i>Streptococcus pneumoniae</i>	
Gram-Negative Bacilli	
<i>Escherichia coli</i>	<i>Serratia</i> spp.
<i>Klebsiella</i> spp.	<i>Acinetobacter</i> spp. ^a
<i>Pseudomonas aeruginosa</i>	<i>Stenotrophomonas</i> spp.
<i>Enterobacter</i> spp.	<i>Citrobacter</i> spp.
<i>Non-aeruginosa Pseudomonas</i> spp. ^a	
Gram-Positive Bacilli	
Diphtheroids	<i>JK bacillus</i> ^a
Fungi	
<i>Candida</i> spp.	<i>Mucor/Rhizopus</i>
<i>Aspergillus</i> spp.	

^aOften associated with intravenous catheters.

received antibiotics for other reasons may develop cellulitis with unusual organisms (e.g., *Escherichia coli*, *Pseudomonas*, or fungi). Early treatment, even of innocent-looking lesions, is essential to prevent necrosis and loss of tissue. Debridement to prevent spread may sometimes be necessary early in the course of disease, but it can often be performed after chemotherapy, when the PMN count increases.

Sweet syndrome, or *febrile neutrophilic dermatosis*, was originally described in women with elevated white blood cell (WBC) counts. The disease is characterized by the presence of leukocytes in the lower dermis, with edema of the papillary body. Ironically, this disease now is usually seen in neutropenic patients with cancer, most often in association with acute myeloid leukemia (AML) but also in association with a variety of other malignancies. Sweet syndrome usually presents as red or bluish-red papules or nodules that may coalesce and form sharply bordered plaques (see Fig. A1-40). The edema may suggest vesicles, but on palpation, the lesions are solid, and vesicles probably never arise in this disease. The lesions are most common on the face, neck, and arms. On the legs, they may be confused with erythema nodosum (see Fig. A1-39). The development of lesions is often accompanied by high fevers and an elevated erythrocyte sedimentation rate. Both the lesions and the temperature elevation respond dramatically to glucocorticoid administration. Treatment begins with high doses of glucocorticoids (prednisone, 60 mg/d) followed by tapered doses over the next 2–3 weeks.

Data indicate that *erythema multiforme* (see Fig. A1-24) with mucous membrane involvement is often associated with herpes simplex virus (HSV) infection and is distinct from Stevens-Johnson syndrome, which is associated with drugs and tends to have a more widespread distribution. Because cancer patients are both immunosuppressed (and therefore susceptible to herpes infections) and heavily treated with drugs (and therefore subject to Stevens-Johnson syndrome [see Fig. A3-4]), both of these conditions are common in this population.

Cytokines, which are used as adjuvants or primary treatments for cancer, can themselves cause characteristic rashes, further complicating the differential diagnosis. This phenomenon is a particular problem in bone marrow (stem cell) transplant recipients (Chap. 143), who, in addition to having the usual chemotherapy-, antibiotic-, and cytokine-induced rashes, are plagued by graft-versus-host disease.

CATHETER-RELATED INFECTIONS

Because intravenous (IV) catheters are commonly used in cancer chemotherapy and are prone to cause infection (Chap. 142), they pose a major problem in the care of patients with cancer. Some catheter-associated infections can be treated with antibiotics, whereas in others, the catheter must be removed (Table 74-5). If the patient has a “tunneled” catheter (which consists of an entrance site, a subcutaneous

tunnel, and an exit site), a red streak over the subcutaneous part of the line (the tunnel) is grounds for immediate device removal. Failure to remove catheters under these circumstances may result in extensive cellulitis and tissue necrosis.

More common than tunnel infections are exit-site infections, often with erythema around the area where the line penetrates the skin. Most authorities (Chap. 147) recommend treatment (usually with vancomycin) for an exit-site infection caused by coagulase-negative *Staphylococcus*. Treatment of coagulase-positive staphylococcal infection is associated with a poorer outcome, and it is advisable to remove the catheter if possible. Similarly, most clinicians remove catheters associated with infections due to *P. aeruginosa* and *Candida* species, because such infections are difficult to treat and bloodstream infections with these organisms are likely to be deadly. Catheter infections caused by *Burkholderia cepacia*, *Stenotrophomonas* species, *Agrobacterium* species, *Acinetobacter baumannii*, *Pseudomonas* species other than *aeruginosa*, and carbapenem-resistant Enterobacteriaceae are likely to be very difficult to eradicate with antibiotics alone. Similarly, isolation of *Bacillus*, *Corynebacterium*, and *Mycobacterium* species should prompt removal of the catheter.

GASTROINTESTINAL TRACT-SPECIFIC SYNDROMES

Upper Gastrointestinal Tract Disease • INFECTIONS OF THE MOUTH The oral cavity is rich in aerobic and anaerobic bacteria (Chap. 177) that normally live in a commensal relationship with the host. The antimetabolic effects of chemotherapy cause a breakdown of mucosal host defenses, leading to ulceration of the mouth and the potential for invasion by resident bacteria. Mouth ulcerations afflict most patients receiving cytotoxic chemotherapy and have been associated with viridans streptococcal bacteremia. *Candida* infections of the mouth are very common. Fluconazole is clearly effective in the treatment of both local infections (thrush) and systemic infections (esophagitis) due to *Candida albicans*. Other azoles (e.g., voriconazole) as well as echinocandins offer similar efficacy as well as activity against the fluconazole-resistant organisms that are associated with chronic fluconazole treatment (Chap. 216).

Noma (*cancrum oris*), commonly seen in malnourished children, is a penetrating disease of the soft and hard tissues of the mouth and adjacent sites, with resulting necrosis and gangrene. It has a counterpart in immunocompromised patients and is thought to be due to invasion of the tissues by *Bacteroides*, *Fusobacterium*, and other normal inhabitants of the mouth. Noma is associated with debility, poor oral hygiene, and immunosuppression.

TABLE 74-5 Approach to Catheter Infections in Immunocompromised Patients

CLINICAL PRESENTATION OR ISOLATED PATHOGEN	CATHETER REMOVAL	ANTIBIOTICS	COMMENTS
Evidence of Infection, Negative Blood Cultures			
Exit-site erythema	Not necessary if infection responds to treatment	Usually, begin treatment for gram-positive cocci.	Coagulase-negative staphylococci are most common.
Tunnel-site erythema	Required	Treat for gram-positive cocci pending culture results.	Failure to remove the catheter may lead to necrosis of the involved area requiring skin grafts in the future.
Blood Culture-Positive Infections			
Coagulase-negative staphylococci	Line removal optimal but may be unnecessary if patient is clinically stable and responds to antibiotics	Usually, start with vancomycin. Linezolid, quinupristin/dalfopristin, and daptomycin are alternative agents.	If there are no contraindications to line removal, this course of action is optimal. If the line is removed, antibiotics may not be necessary.
Other gram-positive cocci (e.g., <i>Staphylococcus aureus</i> , <i>Enterococcus</i>); gram-positive rods (<i>Bacillus</i> , <i>Corynebacterium</i> spp.)	Recommended	Treat with antibiotics to which the organism is sensitive, with duration based on the clinical setting.	The incidence of metastatic infections following <i>S. aureus</i> infection and the difficulty of treating enterococcal infection make line removal the recommended course of action. In addition, gram-positive rods do not respond readily to antibiotics alone.
Gram-negative bacteria	Recommended	Use an agent to which the organism is shown to be sensitive.	Organisms like <i>Stenotrophomonas</i> , <i>Pseudomonas</i> , and <i>Burkholderia</i> are notoriously hard to treat, as are carbapenem-resistant organisms.
Fungi	Recommended	—	Fungal infections of catheters are extremely difficult to treat.

Viruses, particularly HSV, are a prominent cause of morbidity in immunocompromised patients, in whom they are associated with severe mucositis. The use of acyclovir, either prophylactically or therapeutically, is of value.

ESOPHAGEAL INFECTIONS The differential diagnosis of esophagitis (usually presenting as substernal chest pain upon swallowing) includes herpes simplex and candidiasis, both of which are readily treatable.

Lower Gastrointestinal Tract Disease Hepatic candidiasis (*Chap. 216*) results from seeding of the liver (usually from a gastrointestinal source) in neutropenic patients. It is most common among patients being treated for AML and usually presents symptomatically around the time neutropenia resolves. The characteristic picture is that of persistent fever unresponsive to antibiotics, abdominal pain and tenderness or nausea, and elevated serum levels of alkaline phosphatase in a patient with hematologic malignancy who has recently recovered from neutropenia. The diagnosis of this disease (which may present in an indolent manner and persist for several months) is based on the finding of yeasts or pseudohyphae in granulomatous lesions. Hepatic ultrasound or CT may reveal bull's-eye lesions. MRI scans reveal small lesions not visible by other imaging modalities. The pathology (a granulomatous response) and the timing (with resolution of neutropenia and an elevation in granulocyte count) suggest that the host response to *Candida* is an important component of the manifestations of disease. In many cases, although organisms are visible, cultures of biopsied material may be negative. The designation *hepatosplenic candidiasis* or *hepatic candidiasis* is a misnomer because the disease often involves the kidneys and other tissues; the term *chronic disseminated candidiasis* may be more appropriate. Because of the risk of bleeding with liver biopsy, diagnosis is often based on imaging studies (MRI, CT). Treatment should be directed to the causative agent (usually *C. albicans* but sometimes *Candida tropicalis* or other less common *Candida* species).

Typhlitis *Typhlitis* (also referred to as necrotizing colitis, neutropenic colitis, necrotizing enteropathy, ileocecal syndrome, and cecitis) is a clinical syndrome of fever and right-lower-quadrant (or generalized abdominal) tenderness in an immunosuppressed host. This syndrome is classically seen in neutropenic patients after chemotherapy with cytotoxic drugs. It may be more common among children than among adults and appears to be much more common among patients with AML or ALL than among those with other types of cancer. Physical examination reveals right-lower-quadrant tenderness, with or without rebound tenderness. Associated diarrhea (often bloody) is common, and the diagnosis can be confirmed by the finding of a thickened cecal wall on CT, MRI, or ultrasonography. Plain films may reveal a right-lower-quadrant mass, but CT with contrast or MRI is a much more sensitive means of diagnosis. Although surgery is sometimes attempted to avoid perforation from ischemia, most cases resolve with medical therapy alone. The disease is sometimes associated with positive blood cultures (which usually yield aerobic gram-negative bacilli), and therapy is recommended for a broad spectrum of bacteria (particularly gram-negative bacilli, which are likely to be found in the bowel flora).

Clostridioles difficile-Induced Diarrhea Patients with cancer are predisposed to the development of *C. difficile* diarrhea (*Chap. 134*) as a consequence of chemotherapy alone. Thus, they may test positive for *C. difficile* even without receiving antibiotics. Obviously, such patients are also subject to *C. difficile*-induced diarrhea as a result of antibiotic pressure. *C. difficile* should always be considered as a possible cause of diarrhea in cancer patients who have received either chemotherapy or antibiotics. New approaches to treat *C. difficile*-induced diarrhea and to prevent *C. difficile* expansion as part of the gut microbiota may make this disease less troublesome in the future.

CENTRAL NERVOUS SYSTEM-SPECIFIC SYNDROMES

Meningitis The presentation of meningitis in patients with lymphoma or CLL and in patients receiving chemotherapy (particularly with glucocorticoids) for solid tumors suggests a diagnosis of

cryptococcal or listerial infection. As noted previously, splenectomized patients are susceptible to rapid, overwhelming infection with encapsulated bacteria (including *S. pneumoniae*, *H. influenzae*, and *N. meningitidis*). Similarly, patients who are antibody-deficient (e.g., those with CLL, those who have received intensive chemotherapy, or those who have undergone bone marrow [stem cell] transplantation) are likely to have infections caused by these bacteria. Other cancer patients, however, because of their defective cellular immunity, are likely to be infected with other pathogens (Table 74-3). Central nervous system (CNS) tuberculosis should be considered, especially in patients from countries where tuberculosis is highly prevalent in the population.

Encephalitis The spectrum of disease resulting from viral encephalitis is expanded in immunocompromised patients. A predisposition to infections with intracellular organisms similar to those encountered in patients with AIDS (*Chap. 202*) is seen in cancer patients receiving (1) high-dose cytotoxic chemotherapy, (2) chemotherapy affecting T-cell function (e.g., fludarabine), or (3) antibodies that eliminate T cells (e.g., anti-CD3, alemtuzumab, anti-CD52) or cytokine activity (anti-tumor necrosis factor agents or interleukin 1 receptor antagonists). Infection with varicella-zoster virus (VZV) has been associated with encephalitis that may be caused by VZV-related vasculitis. Chronic viral infections may also be associated with dementia and encephalitic presentations. A diagnosis of progressive multifocal leukoencephalopathy (*Chap. 138*) should be considered when a patient who has received chemotherapy (rituximab in particular) presents with dementia (Table 74-6). Other abnormalities of the CNS that may be confused with infection include normal-pressure hydrocephalus and vasculitis resulting from CNS irradiation. It may be possible to differentiate these conditions by MRI.

Brain Masses Mass lesions of the brain most often present as headache with or without fever or neurologic abnormalities. Infections associated with mass lesions may be caused by bacteria (particularly *Nocardia*), fungi (particularly *Cryptococcus* or *Aspergillus*), or parasites (*Toxoplasma*). Epstein-Barr virus (EBV)-associated lymphoma may also present as single—or sometimes multiple—mass lesions of the brain. A biopsy may be required for a definitive diagnosis.

PULMONARY INFECTIONS

Pneumonia (*Chap. 126*) in immunocompromised patients may be difficult to diagnose because conventional methods of diagnosis depend on the presence of neutrophils. Bacterial pneumonia in neutropenic patients may present without purulent sputum—or, in fact, without any sputum at all—and may not produce physical findings suggestive of chest consolidation (rales or egophony).

In granulocytopenic patients with persistent or recurrent fever, the chest x-ray pattern may help to localize an infection and thus to determine which investigative tests and procedures should be undertaken and which therapeutic options should be considered (*Table 74-7*). In this setting, a simple chest x-ray is a screening tool; because the impaired host response results in less evidence of consolidation or infiltration, high-resolution CT is recommended for the diagnosis of pulmonary infections. The difficulties encountered in the management of pulmonary infiltrates relate in part to the difficulties of performing diagnostic

TABLE 74-6 Differential Diagnosis of Central Nervous System Infections in Patients with Cancer

FINDINGS ON CT OR MRI	UNDERLYING PREDISPOSITION	
	PROLONGED NEUTROPENIA	DEFECTS IN CELLULAR IMMUNITY*
Mass lesions	<i>Aspergillus</i> , <i>Nocardia</i> , or <i>Cryptococcus</i> brain abscess	Toxoplasmosis, Epstein-Barr virus lymphoma (rare)
Diffuse encephalitis	Progressive multifocal leukoencephalopathy (JC virus)	Infection with varicella-zoster virus, cytomegalovirus, herpes simplex virus, human herpesvirus type 6, JC virus, <i>Listeria</i>

*High-dose glucocorticoid therapy, cytotoxic chemotherapy.

TABLE 74-7 Differential Diagnosis of Chest Infiltrates in Immunocompromised Patients

INFILTRATE	CAUSE OF PNEUMONIA	
	INFECTIOUS	NONINFECTIOUS
Localized	Bacteria (including <i>Legionella</i> , mycobacteria)	Local hemorrhage or embolism, tumor
Nodular	Fungi (e.g., <i>Aspergillus</i> or <i>Mucor</i> , <i>Nocardia</i>)	Recurrent tumor
Diffuse	Viruses (especially cytomegalovirus), <i>Chlamydia</i> , <i>Pneumocystis</i> , <i>Toxoplasma gondii</i> , mycobacteria	Congestive heart failure, radiation pneumonitis, drug-induced lung injury, lymphangitic spread of cancer

procedures on the patients involved. When platelet counts can be increased to adequate levels by transfusion, microscopic and microbiologic evaluation of the fluid obtained by endoscopic bronchial lavage is often diagnostic. Lavage fluid should be cultured for *Mycoplasma*, *Chlamydia*, *Legionella*, *Nocardia*, more common bacterial pathogens, fungi, and viruses. In addition, the possibility of *Pneumocystis* pneumonia should be considered, especially in patients with ALL or lymphoma who have not received prophylactic trimethoprim-sulfamethoxazole (TMP-SMX). The characteristics of the infiltrate may be helpful in decisions about further diagnostic and therapeutic maneuvers. Nodular infiltrates suggest fungal pneumonia (e.g., that caused by *Aspergillus* or *Mucor*). Such lesions may best be approached by visualized biopsy procedures. It is worth noting that while bacterial pneumonias classically present as lobar infiltrates in normal hosts, bacterial pneumonias in granulocytopenic hosts present with a paucity of signs, symptoms, or radiographic abnormalities; thus, the diagnosis is difficult.

Aspergillus species (Chap. 217) can colonize the skin and respiratory tract or cause fatal systemic illness. Although this fungus may cause aspergillomas in a previously existing cavity or may produce allergic bronchopulmonary disease in some patients, the major problem posed by this genus in neutropenic patients is invasive disease, primarily due to *Aspergillus fumigatus* or *Aspergillus flavus*. The organisms enter the host following colonization of the respiratory tract, with subsequent invasion of blood vessels. The disease is likely to present as a thrombotic or embolic event because of this ability of the fungi to invade blood vessels. The risk of infection with *Aspergillus* correlates directly with the duration of neutropenia. In prolonged neutropenia, positive surveillance cultures for nasopharyngeal colonization with *Aspergillus* may predict the development of disease.

Patients with *Aspergillus* infection often present with pleuritic chest pain and fever, which are sometimes accompanied by cough. Hemoptysis may be an ominous sign. Chest x-rays may reveal new focal infiltrates or nodules. Chest CT may reveal a characteristic halo consisting of a mass-like infiltrate surrounded by an area of low attenuation. The presence of a "crescent sign" on chest x-ray or chest CT, in which the mass progresses to central cavitation, is characteristic of invasive *Aspergillus* infection but may develop as the lesions are resolving.

In addition to causing pulmonary disease, *Aspergillus* may invade through the nose or palate, with deep sinus penetration. The appearance of a discolored area in the nasal passages or on the hard palate should prompt a search for invasive *Aspergillus*. This situation is likely to require surgical debridement. Catheter infections with *Aspergillus* usually require both removal of the catheter and antifungal therapy. Antifungal prophylaxis has led to the emergence of non-fumigatus *Aspergillus* species as well as *Mucorales* and *Scedosporium/Lomentospora* spp. (Chaps. 217–219).

Diffuse interstitial infiltrates suggest viral, parasitic, or *Pneumocystis* pneumonia. If the patient has a diffuse interstitial pattern on chest x-ray, it may be reasonable, while considering invasive diagnostic procedures, to institute empirical treatment for *Pneumocystis* with TMP-SMX and for *Chlamydia*, *Mycoplasma*, and *Legionella* with a quinolone or azithromycin. Noninvasive procedures, such as staining of induced sputum smears for *Pneumocystis*, serum cryptococcal antigen tests, and urine testing for *Legionella* antigen, may be helpful. Serum galactomannan and β -D-glucan tests may be of value in diagnosing *Aspergillus*

infection, but their utility is limited by their lack of sensitivity and specificity. The presence of an elevated level of β -D-glucan in the serum of a patient being treated for cancer who is not receiving prophylaxis against *Pneumocystis* suggests the diagnosis of *Pneumocystis* pneumonia. Infections with viruses that cause only upper respiratory symptoms in immunocompetent hosts, such as respiratory syncytial virus (RSV), influenza viruses, and parainfluenza viruses, may be associated with fatal pneumonitis in immunocompromised hosts. CMV reactivation occurs in cancer patients receiving chemotherapy, but CMV pneumonia is most common among hematopoietic stem cell transplant (HSCT) recipients (Chap. 143). Polymerase chain reaction testing now allows rapid diagnosis of viral pneumonia, which can lead to treatment in some cases (e.g., influenza). Multiplex studies that can detect a wide array of viruses in the lung and upper respiratory tract are now available and will lead to specific diagnoses of viral pneumonias.

Bleomycin is the most common cause of chemotherapy-induced lung disease. Other causes include alkylating agents (such as cyclophosphamide, chlorambucil, and melphalan), nitrosoureas (carmustine [BCNU], lomustine [CCNU], and methyl-CCNU), busulfan, procarbazine, methotrexate, and hydroxyurea. Both infectious and noninfectious (drug- and/or radiation-induced) pneumonitis can cause fever and abnormalities on chest x-ray; thus, the differential diagnosis of an infiltrate in a patient receiving chemotherapy encompasses a broad range of conditions (Table 74-7). The treatment of radiation pneumonitis (which may respond dramatically to glucocorticoids) or drug-induced pneumonitis is different from that of infectious pneumonia, and a biopsy may be important in the diagnosis. Unfortunately, no definitive diagnosis can be made in ~30% of cases, even after bronchoscopy.

Open-lung biopsy is the gold standard of diagnostic techniques. Biopsy via a visualized thoracostomy can replace an open procedure in many cases. When a biopsy cannot be performed, empirical treatment can be undertaken; a quinolone or an erythromycin derivative (azithromycin) and TMP-SMX are used in the case of diffuse infiltrates, and an antifungal agent is administered in the case of nodular infiltrates. The risks should be weighed carefully in these cases. If inappropriate drugs are administered, empirical treatment may prove toxic or ineffective; either of these outcomes may be riskier than biopsy.

■ CARDIOVASCULAR INFECTIONS

Patients with Hodgkin's disease are prone to persistent infections by *Salmonella*, sometimes (and particularly often in elderly patients) affecting a vascular site. The use of IV catheters deliberately lodged in the right atrium is associated with a high incidence of bacterial endocarditis, presumably related to valve damage followed by bacteremia. Nonbacterial thrombotic endocarditis (marantic endocarditis) has been described in association with a variety of malignancies (most often solid tumors) and may follow bone marrow (stem cell) transplantation as well. The presentation of an embolic event with a new cardiac murmur suggests this diagnosis. Blood cultures are negative in this disease of unknown pathogenesis. Infective endocarditis can be a complication of cancer treatment because of the use of IV catheters that lead to bacterial infection. In addition, patients may present with infective endocarditis as an initial presentation of cancer, particularly in the case of gastrointestinal or genitourinary sources.

■ ENDOCRINE SYNDROMES

Infections of the endocrine system have been described in immunocompromised patients. *Candida* infection of the thyroid may be difficult to diagnose during the neutropenic period. It can be defined by indium-labeled WBC scans or gallium scans after neutrophil counts increase. CMV infection can cause adrenalitis with or without resulting adrenal insufficiency. The presentation of a sudden endocrine anomaly in an immunocompromised patient can be a sign of infection in the involved end organ.

■ MUSCULOSKELETAL INFECTIONS

Infection that is a consequence of vascular compromise, resulting in gangrene, can occur when a tumor restricts the blood supply to

muscles, bones, or joints. The process of diagnosis and treatment of such infection is similar to that in normal hosts, with the following caveats:

1. In terms of diagnosis, a lack of physical findings resulting from a lack of granulocytes in the granulocytopenic patient should make the clinician more aggressive in obtaining tissue rather than more willing to rely on physical signs.
2. In terms of therapy, aggressive debridement of infected tissues may be required. However, it is usually difficult to operate on patients who have recently received chemotherapy, both because of a lack of platelets (which results in bleeding complications) and because of a lack of WBCs (which may lead to secondary infection). A blood culture positive for *Clostridium perfringens*—an organism commonly associated with gas gangrene—can have a number of meanings (**Chap. 154**). *Clostridium septicum* bacteraemia is associated with the presence of an underlying malignancy. Bloodstream infections with intestinal organisms such as *Streptococcus bovis* biotype 1 and *C. perfringens* may arise spontaneously from lower gastrointestinal lesions (tumor or polyps); alternatively, these lesions may be harbingers of invasive disease. The clinical setting must be considered in order to define the appropriate treatment for each case.

■ RENAL AND URETERAL INFECTIONS

Infections of the urinary tract are common among patients whose ureteral excretion is compromised (Table 74-1). *Candida*, which has a predilection for the kidney, can invade either from the bloodstream or in a retrograde manner (via the ureters or bladder) in immunocompromised patients. The presence of “fungus balls” or persistent candiduria suggests invasive disease. Persistent funguria (with *Aspergillus* as well as *Candida*) should prompt a search for a nidus of infection in the kidney.

Certain viruses are typically seen only in immunosuppressed patients. BK virus (polyomavirus hominis 1) has been documented in the urine of bone marrow transplant recipients and, like adenovirus, may be associated with hemorrhagic cystitis.

ABNORMALITIES THAT PREDISPOSE TO INFECTION

(Table 74-1)

■ THE LYMPHOID SYSTEM

It is beyond the scope of this chapter to detail how all the immunologic abnormalities that result from cancer or from chemotherapy for cancer lead to infections. Disorders of the immune system are discussed in other sections of this book. As has been noted, patients with antibody deficiency are predisposed to overwhelming infection with encapsulated bacteria (including *S. pneumoniae*, *H. influenzae*, and *N. meningitidis*). Infections that result from the lack of a functional cellular immune system are described in **Chap. 202**. It is worth mentioning, however, that patients undergoing intensive chemotherapy for any form of cancer will have not only defects due to granulocytopenia but also lymphocyte dysfunction, which may be profound. Thus, these patients—especially those receiving glucocorticoid-containing regimens or drugs that inhibit either T-cell activation (calcineurin inhibitors or drugs like fludarabine, which affect lymphocyte function) or cytokine induction—should be given prophylaxis for *Pneumocystis* pneumonia.

Patients receiving treatment that eliminates B cells (e.g., with anti-CD20 antibodies or rituximab) are especially vulnerable to intercurrent viral infections. The incidence of progressive multifocal leukoencephalopathy (caused by JC virus) is elevated among these patients.

■ THE HEMATOPOIETIC SYSTEM

Initial studies in the 1960s revealed a dramatic increase in the incidence of infections (fatal and nonfatal) among cancer patients with a granulocyte count of <500/ μ L. The use of prophylactic antibacterial agents has reduced the number of bacterial infections, but 35–78% of febrile neutropenic patients being treated for hematologic malignancies develop infections at some time during chemotherapy. Aerobic pathogens (both

gram-positive and gram-negative) predominate in all series, but the exact organisms isolated vary from center to center. Infections with anaerobic organisms are uncommon. Geographic patterns affect the types of fungi isolated. Tuberculosis and malaria are common causes of fever in the developing world and may present in this setting as well.

Neutropenic patients are unusually susceptible to infection with a wide variety of bacteria; thus, antibiotic therapy should be initiated promptly to cover likely pathogens if infection is suspected. Indeed, early initiation of antibacterial agents is mandatory to prevent deaths. Like most immunocompromised patients, neutropenic patients are threatened by their own microbial flora, including gram-positive and gram-negative organisms found commonly on the skin and mucous membranes and in the bowel (Table 74-4). Because treatment with narrow-spectrum agents leads to infection with organisms not covered by the antibiotics used, the initial regimen should target all pathogens likely to be the initial causes of bacterial infection in neutropenic hosts. Studies performed in the 1970s suggested that administration of antimicrobial agents should be continued until neutropenia resolves—that is, the granulocyte count is sustained above 500/ μ L for at least 2 days. Recent studies have indicated that it is reasonable to stop antibiotics in patients who are afebrile and stable after 72 hours of treatment (**Fig. 74-2**). Fever may not resolve prior to granulocyte recovery. In some cases, patients remain febrile after resolution of neutropenia. In these instances, the risk of sudden death from overwhelming bacteremia is greatly reduced, and the following diagnoses should be seriously considered: (1) fungal infection, (2) bacterial abscesses or undrained foci of infection, and (3) drug fever (including reactions to antimicrobial agents as well as to chemotherapy or cytokines). In the proper setting, viral infection or graft-versus-host disease should be considered. In clinical practice, antibacterial therapy is usually discontinued when the patient is no longer neutropenic and all evidence of bacterial disease has been eliminated. Antifungal agents are then discontinued if there is no evidence of fungal disease. If the patient remains febrile, a search for viral diseases or unusual pathogens is conducted while unnecessary cytokines and other drugs are systematically eliminated from the regimen.

TREATMENT

Infections in Cancer Patients

ANTIBACTERIAL THERAPY

Hundreds of antibacterial regimens have been tested for use in patients with cancer. The major risk of infection is related to the degree of neutropenia seen as a consequence of either the disease or the therapy. Many of the relevant studies have involved small populations in which the outcomes have generally been good, and most have lacked the statistical power to detect differences among the regimens studied. Each febrile neutropenic patient should be approached as a unique problem, with particular attention given to previous infections and recent antibiotic exposures. Several general guidelines are useful in the initial treatment of neutropenic patients with fever (**Fig. 74-2**):

1. In the initial regimen, it is necessary to use antibiotics active against both gram-negative and gram-positive bacteria (Table 74-4).
2. Monotherapy with an aminoglycoside or an antibiotic lacking good activity against gram-positive organisms (e.g., ciprofloxacin or aztreonam) is not adequate in this setting.
3. The agents used should reflect both the epidemiology and the antibiotic resistance pattern of the hospital.
4. If the pattern of resistance justifies its use, a single third-generation cephalosporin constitutes an appropriate initial regimen in many hospitals.
5. Most standard regimens are designed for patients who have not previously received prophylactic antibiotics. The development of fever in a patient who has received antibiotics affects the choice of subsequent therapy, which should target resistant organisms and organisms known to cause infections in patients being treated with the antibiotics already administered.

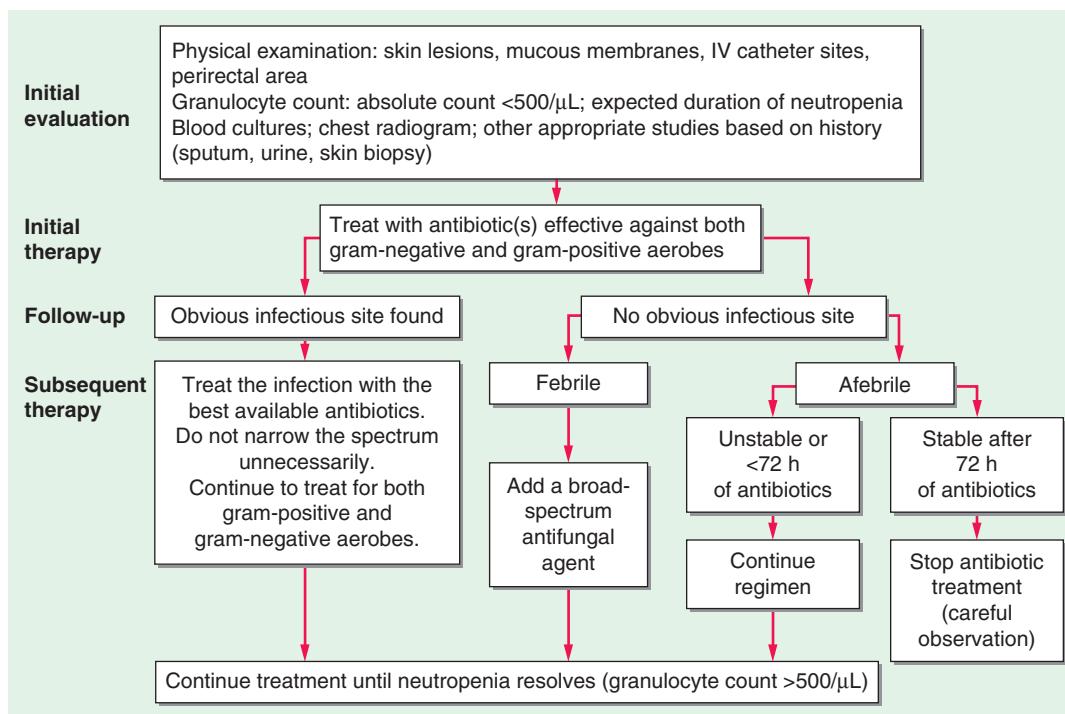


FIGURE 74-2 Algorithm for the diagnosis and treatment of fever and neutropenia.

- Randomized trials have indicated the safety of oral antibiotic regimens in the treatment of “low-risk” patients with fever and neutropenia. Outpatients who are expected to remain neutropenic for <10 days and who do not have concurrent medical problems (such as hypotension, pulmonary compromise, or abdominal pain) can be classified as low risk and treated with a broad-spectrum oral regimen.
- Several large-scale studies indicate that prophylaxis with a fluoroquinolone (ciprofloxacin or levofloxacin) decreases morbidity and mortality rates among afebrile patients who are anticipated to have neutropenia of long duration.

Commonly used antibiotic regimens for the treatment of febrile patients in whom prolonged neutropenia (>7 days) is anticipated include (1) ceftazidime or cefepime, (2) piperacillin/tazobactam, or (3) imipenem/cilastatin or meropenem. All three regimens have shown equal efficacy in large trials. All three are active against *P. aeruginosa* and a broad spectrum of aerobic gram-positive and gram-negative organisms. Imipenem/cilastatin has been associated with an elevated rate of *C. difficile* diarrhea, and many centers reserve carbapenem antibiotics for treatment of gram-negative bacteria that produce extended-spectrum β-lactamases; these limitations make carbapenems less attractive as an initial regimen. Despite the frequent involvement of coagulase-negative staphylococci, the initial use of vancomycin or its automatic addition to the initial regimen has not resulted in improved outcomes, and the antibiotic does exert toxic effects. For these reasons, only judicious use of vancomycin is recommended—for example, when there is good reason to suspect the involvement of coagulase-negative staphylococci (e.g., the appearance of erythema at the exit site of a catheter or a positive culture for methicillin-resistant *S. aureus* or coagulase-negative staphylococci). Because the sensitivities of bacteria vary from hospital to hospital, clinicians are advised to check their local sensitivities and to be aware that resistance patterns can change quickly, necessitating a change in the approach to patients with fever and neutropenia. Similarly, infection control services should monitor for basic antibiotic resistance and for fungal infections. The appearance of a large number of *Aspergillus* infections, in particular, suggests the possibility of an environmental source that requires further investigation and remediation.

The initial antibacterial regimen should be refined on the basis of culture results (Fig. 74-2). Blood cultures are the most relevant basis for selection of therapy; surface cultures of skin and mucous membranes may be misleading. In the case of gram-positive bacteremia or another gram-positive infection, it is important that the antibiotic be optimal for the organism isolated. Once treatment with broad-spectrum antibiotics has begun, it is not desirable to discontinue all antibiotics because of the risk of failing to treat a potentially fatal bacterial infection; the addition of more and more antibacterial agents to the regimen is not appropriate unless there is a clinical or microbiologic reason to do so. Planned progressive therapy (the serial, empirical addition of one drug after another without culture data) is not efficacious in most settings and may have unfortunate consequences. Simply adding another antibiotic for fear that a gram-negative infection is present is a dubious practice. The synergy exhibited by β-lactams and aminoglycosides against certain gram-negative organisms (especially *P. aeruginosa*) provides the rationale for using two antibiotics in this setting, but recent analyses suggest that efficacy is not enhanced by the addition of aminoglycosides, while toxicity may be increased. Mere “double coverage,” with the addition of a quinolone or another antibiotic that is not likely to exhibit synergy, has not been shown to be beneficial and may cause additional toxicities and side effects. Cephalosporins can cause bone marrow suppression, and vancomycin is associated with neutropenia in some healthy individuals. Furthermore, the addition of multiple cephalosporins may induce β-lactamase production by some organisms; cephalosporins and double β-lactam combinations should probably be avoided altogether in *Enterobacter* infections.

ANTIFUNGAL THERAPY

Fungal infections in cancer patients are most often associated with neutropenia. Neutropenic patients are predisposed to the development of invasive fungal infections, most commonly those due to *Candida* and *Aspergillus* species and occasionally those caused by *Mucor*, *Rhizopus*, *Fusarium*, *Trichosporon*, *Bipolaris*, and others. Invasive candidal disease is usually caused by *C. albicans* or *C. tropicalis* but can be caused by *C. krusei*, *C. parapsilosis*, and *C. glabrata*. The worldwide spread of *C. auris*, a species that is typically resistant to fluconazole and often resistant to amphotericin B as

well, has further complicated the management of invasive *Candida* infections ([Chap. 216](#)).

For decades, it has been common clinical practice to add amphotericin B to antibacterial regimens if a neutropenic patient remains febrile despite 4–7 days of treatment with antibacterial agents. The rationale for this empirical addition is that it is difficult to culture fungi before they cause disseminated disease and that mortality rates from disseminated fungal infections in granulocytopenic patients are high. Before the introduction of newer azoles into clinical practice, amphotericin B was the mainstay of antifungal therapy. The insolubility of amphotericin B has resulted in the marketing of several lipid formulations that are less toxic than the amphotericin B deoxycholate complex. Echinocandins (e.g., caspofungin) are useful in the treatment of infections caused by azole-resistant *Candida* strains as well as in therapy for aspergillosis and have been shown to be equivalent to liposomal amphotericin B for the empirical treatment of patients with prolonged fever and neutropenia. Newer azoles have also been demonstrated to be effective in this setting. Although fluconazole is efficacious in the treatment of infections due to many *Candida* species, its use against serious fungal infections in immunocompromised patients is limited by its narrow spectrum: it has no activity against *Aspergillus* or against several non-*albicans* *Candida* species. The broad-spectrum azoles (e.g., voriconazole and posaconazole) provide another option for the treatment of *Aspergillus* infections ([Chap. 217](#)), including CNS infection. Clinicians should be aware that the spectrum of each azole is somewhat different and that no drug can be assumed to be efficacious against all fungi. *Aspergillus terreus* is resistant to amphotericin B. Although voriconazole is active against *Pseudallescheria boydii*, amphotericin B is not; however, voriconazole has no activity against *Mucor*. Posaconazole, which is administered orally, is useful as a prophylactic agent in patients with prolonged neutropenia. Studies in progress are assessing the use of these agents in combinations. For a full discussion of antifungal therapy, see [Chap. 211](#).

ANTIVIRAL THERAPY

The availability of a variety of agents active against herpes-group viruses, including some new agents with a broader spectrum of activity, has heightened focus on the treatment of viral infections, which pose a major problem in cancer patients. Viral diseases caused by the herpes group are prominent. Serious (and sometimes fatal) infections due to HSV and VZV are well documented in patients receiving chemotherapy. CMV may also cause serious disease, but fatalities from CMV infection are more common in hematopoietic stem cell transplant recipients. The roles of human herpesvirus (HHV)-6, HHV-7, and HHV-8 (Kaposi's sarcoma-associated herpesvirus) in cancer patients are still being defined ([Chap. 195](#)). EBV lymphoproliferative disease (LPD) can occur in patients receiving chemotherapy but is much more common among transplant recipients ([Chap. 143](#)). While clinical experience is most extensive with acyclovir, which can be used therapeutically or prophylactically, a number of derivative drugs offer advantages over this agent ([Chap. 191](#)).

In addition to the herpes group, several respiratory viruses (especially RSV) may cause serious disease in cancer patients. Although influenza vaccination is recommended (see below), it may be ineffective in this patient population. The availability of antiviral drugs with activity against influenza viruses gives the clinician additional options for the prophylaxis and treatment of these patients ([Chaps. 191 and 200](#)).

The COVID-19 pandemic has affected cancer patients disproportionately. Early analyses suggest that lung cancer patients in particular are more vulnerable to serious infection with SARS-CoV-2.

OTHER THERAPEUTIC MODALITIES

A variety of cytokines, including granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor, enhance granulocyte recovery after chemotherapy and consequently shorten the period of maximal vulnerability to fatal infections. Most authorities recommend their use only when neutropenia is both severe and prolonged, and they should be used only in the

appropriate setting (i.e., when stem cells are likely to be responsive) and not as an adjunct to antimicrobial agents. The cytokines themselves may have adverse effects, including fever, hypoxemia, and pleural effusions or serositis in other areas ([Chap. 349](#)).

Once neutropenia has resolved, the risk of infection decreases dramatically. However, depending on what drugs they receive, patients who continue on chemotherapeutic protocols remain at high risk for certain diseases. Any patient receiving more than a maintenance dose of glucocorticoids (e.g., in many treatment regimens for diffuse lymphoma) should also receive prophylactic TMP-SMX because of the risk of *Pneumocystis* infection; those with ALL should receive such prophylaxis for the duration of chemotherapy.

PREVENTION OF INFECTION IN CANCER PATIENTS

EFFECT OF THE ENVIRONMENT

Outbreaks of fatal *Aspergillus* infection have been associated with construction projects and materials in several hospitals. The association between spore counts and risk of infection suggests the need for a high-efficiency air-handling system in hospitals that care for large numbers of neutropenic patients. The use of laminar-flow rooms and prophylactic antibiotics has decreased the number of infectious episodes in severely neutropenic patients. However, because of the expense of such a program and the failure to show that it dramatically affects mortality rates, most centers do not routinely use laminar flow to care for neutropenic patients. Some centers use “reverse isolation,” in which health care providers and visitors to a patient who is neutropenic wear gowns and gloves. Since most of the infections these patients develop are due to organisms that colonize the patients' own skin and bowel, the validity of such schemes is dubious, and limited clinical data do not support their use. Hand washing by all staff caring for neutropenic patients should be required to prevent the spread of resistant organisms.

The presence of large numbers of bacteria (particularly *P. aeruginosa*) in certain foods, especially fresh vegetables, has led some authorities to recommend a special “low-bacteria” diet. A diet consisting of cooked and canned food is satisfactory to most neutropenic patients and does not involve elaborate disinfection or sterilization protocols. However, there are no studies to support even this type of dietary restriction. Counseling of patients to avoid leftovers, deli foods, undercooked meat, and unpasteurized dairy products is recommended since these foods have been associated with outbreaks of listerial infection.

PHYSICAL MEASURES

Although few studies address this issue, patients with cancer are predisposed to infections resulting from anatomic compromise (e.g., lymphedema resulting from node dissections after radical mastectomy). Surgeons who specialize in cancer surgery can provide specific guidelines for the care of such patients, and patients benefit from common-sense advice about how to prevent infections in vulnerable areas.

IMMUNOGLOBULIN REPLACEMENT

Many patients with multiple myeloma or CLL have immunoglobulin deficiencies as a result of their disease, and all allogeneic bone marrow transplant recipients are hypogammaglobulinemic for a period after transplantation. However, current recommendations reserve intravenous immunoglobulin replacement therapy for patients with severe, prolonged hypogammaglobulinemia (<400 mg of total IgG/dL) and a history of repeated infections. Antibiotic prophylaxis has been shown to be cheaper and is efficacious in preventing infections in most CLL patients with hypogammaglobulinemia. Routine use of immunoglobulin replacement is not recommended.

SEXUAL PRACTICES

The use of condoms is recommended for severely immunocompromised patients. Any sexual practice that results in oral exposure to feces is not recommended. Neutropenic patients should be advised to avoid any practice that results in trauma, as even microscopic cuts may result in bacterial invasion and fatal sepsis.

■ ANTIBIOTIC PROPHYLAXIS

Several studies indicate that the use of oral fluoroquinolones prevents infection and decreases mortality rates among severely neutropenic patients. Prophylaxis for *Pneumocystis* is mandatory for patients with ALL and for all cancer patients receiving glucocorticoid-containing chemotherapy regimens.

■ VACCINATION OF CANCER PATIENTS

In general, patients undergoing chemotherapy respond less well to vaccines than do normal hosts. Their greater need for vaccines thus leads to a dilemma in their management. Purified proteins and inactivated vaccines are almost never contraindicated and should be given to patients even during chemotherapy. For example, all adults should receive diphtheria–tetanus toxoid boosters at the indicated times as well as seasonal influenza vaccine. However, if possible, vaccination should not be undertaken concurrent with cytotoxic chemotherapy. If patients are expected to be receiving chemotherapy for several months and vaccination is indicated (e.g., influenza vaccination in the fall), the vaccine should be given midcycle—as far apart in time as possible from the antimetabolic agents that will prevent an immune response. The meningococcal and pneumococcal polysaccharide vaccines should be given to patients before splenectomy, if possible. The *H. influenzae* type b conjugate vaccine should be administered to all splenectomized patients.

In general, live virus (or live bacterial) vaccines should not be given to patients during intensive chemotherapy because of the risk of disseminated infection. Recommendations on vaccination are summarized in Table 74-2 (see <https://www.cdc.gov/vaccines/hcp/index.html> for updated recommendations).

■ IN MEMORIAM

Dr. Robert W. Finberg, Richard M. Haidack Distinguished Professor and Chair of Medicine, University of Massachusetts Chan Medical School (2000-2020), Professor of Medicine and Chair of Infectious Diseases, Dana Farber Cancer Institute (1996-1999), passed away on August 30, 2021. In addition to this chapter, he authored Chapter 143, “Infections in Transplant Recipients.” Dr. Finberg was an internationally renowned physician-scientist and an academic leader whose career spanned four decades. A brilliant talented researcher focused on viral pathogenesis he was also a consummate clinician who attended at the bedside throughout his career. Dr. Finberg played an important role in the COVID-19 pandemic by leading clinical trials for SARS-CoV-2 vaccines and therapeutics. Warm and generous with a keen wit, he was a beloved family man, colleague and friend. As an educator and mentor he truly cared about training the next generation, as evidenced by the legacy of a very large number of trainees he leaves behind. We are indebted to Dr. Finberg for his outstanding contributions to nine editions of Harrison’s Principles of Internal Medicine and to his considerable and significant contributions to the field of human health.

■ FURTHER READING

- FERNÁNDEZ-CRUZ A et al: Infective endocarditis in patients with cancer: A consequence of invasive procedures or a harbinger of neoplasm? A prospective, multicenter cohort. *Medicine* 96:e7913, 2017.
- FRIEDMAN DZP, SCHWARTZ IS: Emerging fungal infections: New patients, new patterns, and new pathogens. *J Fungi* 5:67, 2019.
- MASCHMEYER G et al: Infections associated with immunotherapeutic and molecular targeted agents in hematology and oncology. A position paper by the European Conference on Infections in Leukemia (ECIL). *Leukemia* 33:844, 2019.
- PIZZO PA: Management of patients with fever and neutropenia through the arc of time. *Ann Intern Med* 170:389, 2019.
- ZHANG L et al: Clinical characteristics of COVID-19-infected cancer patients: A retrospective case study in three hospitals within Wuhan, China. *Ann Oncol* 31:894, 2020.

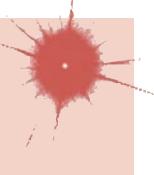
■ WEBSITE

Prevention and Treatment of Cancer-Related Infections; National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology Version 2.2020 (<https://www.nccn.org.org>).

75

Oncologic Emergencies

Rasim Gucalp, Janice P. Dutcher



Emergencies in patients with cancer may be classified into three groups: pressure or obstruction caused by a space-occupying lesion, metabolic or hormonal problems (paraneoplastic syndromes, [Chap. 93](#)), and treatment-related complications.

STRUCTURAL-OBSTRUCTIVE ONCOLOGIC EMERGENCIES

■ SUPERIOR VENA CAVA SYNDROME

Superior vena cava syndrome (SVCS) is the clinical manifestation of superior vena cava (SVC) obstruction, with severe reduction in venous return from the head, neck, and upper extremities. Malignant tumors, such as lung cancer, lymphoma, and metastatic tumors, are responsible for the majority of SVCS cases. With the expanding use of intravascular devices (e.g., permanent central venous access catheters, pacemaker/defibrillator leads), the prevalence of benign causes of SVCS is now increasing, accounting for at least 40% of cases. Lung cancer, particularly of small-cell and squamous cell histologies, accounts for ~85% of all cases of malignant origin. In young adults, malignant lymphoma is a leading cause of SVCS. Hodgkin’s lymphoma involves the mediastinum more commonly than other lymphomas but rarely causes SVCS. When SVCS is noted in a young man with a mediastinal mass, the differential diagnosis is lymphoma versus primary mediastinal germ cell tumor. Metastatic cancers to the mediastinal lymph nodes, such as testicular and breast carcinomas, account for a small proportion of cases. Other causes include benign tumors, aortic aneurysm, thyromegaly, thrombosis, and fibrosing mediastinitis from prior irradiation, histoplasmosis, or Behcet’s syndrome. SVCS as the initial manifestation of Behcet’s syndrome may be due to inflammation of the SVC associated with thrombosis.

Patients with SVCS usually present with neck and facial swelling (especially around the eyes), dyspnea, and cough. Other symptoms include hoarseness, tongue swelling, headaches, nasal congestion, epistaxis, hemoptysis, dysphagia, pain, dizziness, syncope, and lethargy. Bending forward or lying down may aggravate the symptoms. The characteristic physical findings are dilated neck veins; an increased number of collateral veins covering the anterior chest wall; cyanosis; and edema of the face, arms, and chest. Facial swelling and plethora are typically exacerbated when the patient is supine. More severe cases include proptosis, glossal and laryngeal edema, and obtundation. The clinical picture is milder if the obstruction is located above the azygos vein. Symptoms are usually progressive, but in some cases, they may improve as collateral circulation develops.

Signs and symptoms of cerebral and/or laryngeal edema, though rare, are associated with a poorer prognosis and require urgent evaluation. Seizures are more likely related to brain metastases than to cerebral edema from venous occlusion. Patients with small-cell lung cancer and SVCS have a higher incidence of brain metastases than those without SVCS.

Cardiorespiratory symptoms at rest, particularly with positional changes, suggest significant airway and vascular obstruction and limited physiologic reserve. Cardiac arrest or respiratory failure can occur, particularly in patients receiving sedatives or undergoing general anesthesia.

Rarely, esophageal varices may develop, particularly in the setting of SVC syndrome due to hemodialysis catheter. These are “downhill” varices based on the direction of blood flow from cephalad to caudad (in contrast to “uphill” varices associated with caudad to cephalad flow from portal hypertension). If the obstruction to the SVC is proximal to the azygous vein, varices develop in the upper one-third of the esophagus. If the obstruction involves or is distal to the azygous vein, varices occur in the entire length of the esophagus. Variceal bleeding may be a late complication of chronic SVCS.

SVC obstruction may lead to bilateral breast edema with bilateral enlarged breasts. Unilateral breast dilation may be seen as a consequence of axillary or subclavian vein blockage.

The diagnosis of SVCS is a clinical one. The most significant chest radiographic finding is widening of the superior mediastinum, most commonly on the right side. Pleural effusion occurs in only 25% of patients, often on the right side. The majority of these effusions are exudative and occasionally chylous. However, a normal chest radiograph is still compatible with the diagnosis if other characteristic findings are present. Computed tomography (CT) provides the most reliable view of the mediastinal anatomy. The diagnosis of SVCS requires diminished or absent opacification of central venous structures with prominent collateral venous circulation. Magnetic resonance imaging (MRI) is increasingly being used to diagnose SVC obstruction with a 100% sensitivity and specificity, but dyspneic SVCS patients may have difficulty remaining supine for the entire imaging process. Invasive procedures, including bronchoscopy, percutaneous needle biopsy, mediastinoscopy, and even thoracotomy, can be performed by a skilled clinician without any major risk of bleeding. Endobronchial or esophageal ultrasound-guided needle aspiration may establish the diagnosis safely. For patients with a known cancer, a detailed workup usually is not necessary, and appropriate treatment may be started after obtaining a CT scan of the thorax. For those with no history of malignancy, a detailed evaluation is essential to rule out benign causes and determine a specific diagnosis to direct the appropriate therapy.

TREATMENT

Superior Vena Cava Syndrome

The one potentially life-threatening complication of a superior mediastinal mass is tracheal obstruction. Upper airway obstruction demands emergent therapy. Diuretics with a low-salt diet, head elevation, and oxygen may produce temporary symptomatic relief. Glucocorticoids have a limited role except in the setting of mediastinal lymphoma masses.

Radiation therapy is the primary treatment for SVCS caused by non-small-cell lung cancer and other metastatic solid tumors. Chemotherapy is effective when the underlying cancer is small-cell carcinoma of the lung, lymphoma, or germ cell tumor. SVCS recurs in 10–30% of patients; it may be palliated with the use of intravascular self-expanding stents (Fig. 75-1). Endovascular therapy is more frequently used first, to provide rapid relief of clinical symptoms with reduced complications. Early stenting may be necessary in patients with severe symptoms; however, the prompt increase in venous return after stenting may precipitate heart failure and pulmonary edema. Other complications of stent placement include hematoma at the insertion site, SVC perforation, stent migration in the right ventricle, stent fracture, and pulmonary embolism.

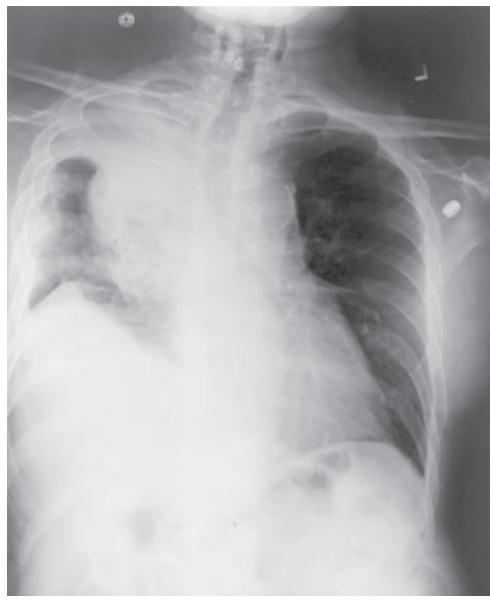
Clinical improvement occurs in most patients, although this improvement may be due to the development of adequate collateral circulation. The mortality associated with SVCS does not relate to caval obstruction but rather to the underlying cause.

SVCS AND CENTRAL VENOUS CATHETERS IN ADULTS

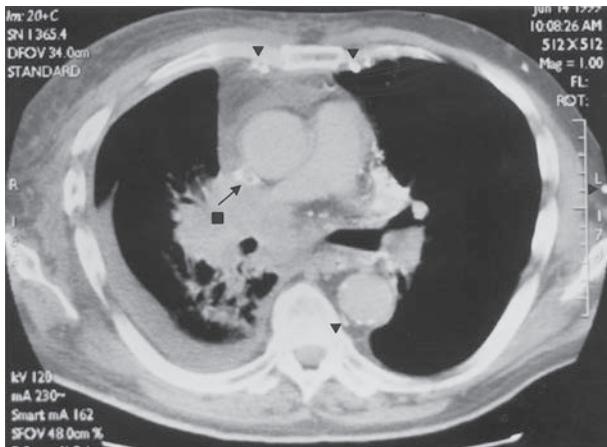
The use of long-term central venous catheters has become common practice in patients with cancer. Major vessel thrombosis may occur. In these cases, catheter removal should be combined with anticoagulation to prevent embolization. SVCS in this setting, if detected early, can be treated by fibrinolytic therapy without sacrificing the catheter. When managing patients with transvenous lead-related SVC syndrome, anticoagulation, local and systemic thrombolytic therapy, and surgical intervention can be effective therapy in select patients. Endovascular stenting has also been shown to be safe and promising, with minimal procedural or clinical complications. The role of anticoagulation after SVC stent placement is controversial.

PERICARDIAL EFFUSION/TAMPOONADE

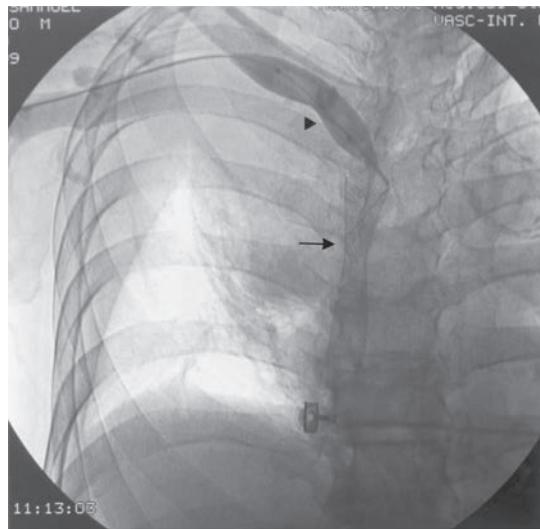
Malignant pericardial disease is found at autopsy in 5–10% of patients with cancer, most frequently with lung cancer, breast cancer, leukemias,



A



B



C

FIGURE 75-1 Superior vena cava syndrome (SVCS). **A.** Chest radiographs of a 59-year-old man with recurrent SVCS caused by non-small-cell lung cancer showing right paratracheal mass with right pleural effusion. **B.** Computed tomography of same patient demonstrating obstruction of the superior vena cava with thrombosis (arrow) by the lung cancer (square) and collaterals (arrowheads). **C.** Balloon angioplasty (arrowhead) with Wallstent (arrow) in same patient.

and lymphomas. Cardiac tamponade as the initial presentation of extrathoracic malignancy is rare. The origin is not malignancy in ~50% of cancer patients with symptomatic pericardial disease, but it can be related to irradiation; drug-induced pericarditis, including chemotherapeutic agents such as all-trans retinoic acid, arsenic trioxide, imatinib, and other abl kinase inhibitors; hypothyroidism; idiopathic pericarditis; infection; or autoimmune diseases. Pericardial disease has been associated with immune checkpoint inhibitors specifically in patients with advanced non-small-cell lung cancer. Two types of radiation pericarditis occur: an acute inflammatory, effusive pericarditis occurring within months of irradiation, which usually resolves spontaneously, and a chronic effusive pericarditis that may appear up to 20 years after radiation therapy and is accompanied by a thickened pericardium.

Most patients with pericardial metastasis are asymptomatic. However, the common symptoms are dyspnea, cough, chest pain, orthopnea, and weakness. Pleural effusion, sinus tachycardia, jugular venous distention, hepatomegaly, peripheral edema, and cyanosis are the most frequent physical findings. Relatively specific diagnostic findings, such as paradoxical pulse, diminished heart sounds, pulsus alternans (pulse waves alternating between those of greater and lesser amplitude with successive beats), and friction rub are less common than with non-malignant pericardial disease. Chest radiographs and electrocardiogram (ECG) reveal abnormalities in 90% of patients, but half of these abnormalities are nonspecific. Echocardiography is the most helpful diagnostic test. Pericardial fluid may be serous, serosanguineous, or hemorrhagic, and cytologic examination of pericardial fluid is diagnostic in most patients. Measurements of tumor markers in the pericardial fluid are not helpful in the diagnosis of malignant pericardial fluid. Pericardioscopy with targeted pericardial and epicardial biopsy may differentiate neoplastic and benign pericardial disease. A combination of cytology, pericardial and epicardial biopsy, and guided pericardioscopy gives the best diagnostic yield. CT scan of chest may also reveal the presence of a concomitant thoracic neoplasm. Cancer patients with pericardial effusion containing malignant cells on cytology have a very poor survival, ~7 weeks.

TREATMENT

Pericardial Effusion/Tamponade

Pericardiocentesis with or without the introduction of sclerosing agents, the creation of a pericardial window, complete pericardial stripping, cardiac irradiation, or systemic chemotherapy are effective treatments. Acute pericardial tamponade with life-threatening hemodynamic instability requires immediate drainage of fluid. This can be quickly achieved by pericardiocentesis. The recurrence rate after percutaneous catheter drainage is ~20%. Sclerotherapy (pericardial instillation of bleomycin, mitomycin C, or tetracycline) may decrease recurrences. Alternatively, subxiphoid pericardiotomy can be performed in 45 min under local anesthesia. Thoracoscopic pericardial fenestration can be employed for benign causes; however, 60% of malignant pericardial effusions recur after this procedure. In a subset of patients, drainage of the pericardial effusion is paradoxically followed by worsening hemodynamic instability. This so-called “postoperative low cardiac output syndrome” occurs in up to 10% of patients undergoing surgical drainage and carries poor short-term survival.

■ INTESTINAL OBSTRUCTION

Intestinal obstruction and reobstruction are common problems in patients with advanced cancer, particularly colorectal or ovarian carcinoma. However, other cancers, such as lung or breast cancer and melanoma, can metastasize within the abdomen, leading to intestinal obstruction. Metastatic disease from colorectal, ovarian, pancreatic, gastric, and occasionally breast cancer can lead to peritoneal carcinomatosis, with infiltration of the omentum and peritoneal surface, thus limiting bowel motility. Typically, obstruction occurs at multiple sites in peritoneal carcinomatosis. Melanoma has a predilection to involve the small bowel; this involvement may be isolated, and resection may

result in prolonged survival. Intestinal pseudoobstruction is caused by infiltration of the mesentery or bowel muscle by tumor, involvement of the celiac plexus, or paraneoplastic neuropathy in patients with small-cell lung cancer. Paraneoplastic neuropathy is associated with IgG antibodies reactive to neurons of the myenteric and submucosal plexuses of the jejunum and stomach. Ovarian cancer can lead to authentic luminal obstruction or to pseudoobstruction that results when circumferential invasion of a bowel segment arrests the forward progression of peristaltic contractions.

The onset of obstruction is usually insidious. Pain is the most common symptom and is usually colicky in nature. Pain can also be due to abdominal distention, tumor masses, or hepatomegaly. Vomiting can be intermittent or continuous. Patients with complete obstruction usually have constipation. Physical examination may reveal abdominal distention with tympany, ascites, visible peristalsis, high-pitched bowel sounds, and tumor masses. Erect plain abdominal films may reveal multiple air-fluid levels and dilation of the small or large bowel. Acute cecal dilation to >12–14 cm is considered a surgical emergency because of the high likelihood of rupture. CT scan is useful in defining the extent of disease and the exact nature of the obstruction and differentiating benign from malignant causes of obstruction in patients who have undergone surgery for malignancy. Malignant obstruction is suggested by a mass at the site of obstruction or prior surgery, adenopathy, or an abrupt transition zone and irregular bowel thickening at the obstruction site. Benign obstruction is more likely when CT shows mesenteric vascular changes, a large volume of ascites, or a smooth transition zone and smooth bowel thickening at the obstruction site. In challenging patients with obstructive symptoms, particularly low-grade small-bowel obstruction (SBO), CT enteroclysis often can help establish the diagnosis by providing distention of small-bowel loops. In this technique, water-soluble contrast is infused through a nasoenteric tube into the duodenum or proximal small bowel followed by CT images. The prognosis for the patient with cancer who develops intestinal obstruction is poor; median survival is 3–4 months. About 25–30% of patients are found to have intestinal obstruction due to causes other than cancer. Adhesions from previous operations are a common benign cause. Ileus induced by vinca alkaloids, narcotics, or other drugs is another reversible cause.

TREATMENT

Intestinal Obstruction

The management of intestinal obstruction in patients with advanced malignancy depends on the extent of the underlying malignancy, options for further antineoplastic therapy, estimated life expectancy, the functional status of the major organs, and the extent of the obstruction. The initial management should include surgical evaluation. Operation is not always successful and may lead to further complications with a substantial mortality rate (10–20%). Laparoscopy can diagnose and treat malignant bowel obstruction in some cases. Self-expanding metal stents placed in the gastric outlet, duodenum, proximal jejunum, colon, or rectum may palliate obstructive symptoms at those sites without major surgery. Patients known to have advanced intraabdominal malignancy should receive a prolonged course of conservative management, including nasogastric decompression. Percutaneous endoscopic or surgical gastrostomy tube placement is an option for palliation of nausea and vomiting, the so-called “venting gastrostomy.” Treatment with antiemetics, antispasmodics, and analgesics may allow patients to remain outside the hospital. Octreotide may relieve obstructive symptoms through its inhibitory effect on gastrointestinal secretion. Glucocorticoids have anti-inflammatory effects and may help the resolution of bowel obstruction. They also have antiemetic effects.

■ URINARY OBSTRUCTION

Urinary obstruction may occur in patients with prostatic or gynecologic malignancies, particularly cervical carcinoma; metastatic disease

from other primary sites such as carcinomas of the breast, stomach, lung, colon, and pancreas; or lymphomas. Radiation therapy to pelvic tumors may cause fibrosis and subsequent ureteral obstruction. Bladder outlet obstruction is usually due to prostate and cervical cancers and may lead to bilateral hydronephrosis and renal failure.

Flank pain is the most common symptom. Persistent urinary tract infection, persistent proteinuria, or hematuria in patients with cancer should raise suspicion of ureteral obstruction. Total anuria and/or anuria alternating with polyuria may occur. A slow, continuous rise in the serum creatinine level necessitates immediate evaluation. Renal ultrasound is the safest and cheapest way to identify hydronephrosis. The function of an obstructed kidney can be evaluated by a nuclear scan. CT scan can reveal the point of obstruction and identify a retroperitoneal mass or adenopathy.

TREATMENT

Urinary Obstruction

Obstruction associated with flank pain, sepsis, or fistula formation is an indication for immediate palliative urinary diversion. Internal ureteral stents can be placed under local anesthesia. Percutaneous nephrostomy offers an alternative approach for drainage. The placement of a nephrostomy is associated with a significant rate of pyelonephritis. In the case of bladder outlet obstruction due to malignancy, a suprapubic cystostomy can be used for urinary drainage. An aggressive intervention with invasive approaches to improve the obstruction should be weighed against the likelihood of antitumor response, and the ability to reverse renal insufficiency should be evaluated.

MALIGNANT BILIARY OBSTRUCTION

This common clinical problem can be caused by a primary carcinoma arising in the pancreas, ampulla of Vater, bile duct, or liver or by metastatic disease to the periductal lymph nodes or liver parenchyma. The most common metastatic tumors causing biliary obstruction are gastric, colon, breast, and lung cancers. Jaundice, light-colored stools, dark urine, pruritus, and weight loss due to malabsorption are usual symptoms. Pain and secondary infection are uncommon in malignant biliary obstruction. Ultrasound, CT scan, or percutaneous transhepatic or endoscopic retrograde cholangiography will identify the site and nature of the biliary obstruction.

TREATMENT

Malignant Biliary Obstruction

Palliative intervention is indicated only in patients with disabling pruritus resistant to medical treatment, severe malabsorption, or infection. Stenting under radiographic control, surgical bypass, or radiation therapy with or without chemotherapy may alleviate the obstruction. The choice of therapy should be based on the site of obstruction (proximal vs distal), the type of tumor (sensitive to radiotherapy, chemotherapy, or neither), and the general condition of the patient. Stenting under radiographic or endoscopic control, surgical bypass, or radiation therapy with or without chemotherapy may alleviate the obstruction. Photodynamic therapy and radiofrequency ablation are promising endoscopic therapies for malignant biliary obstruction.

Endoscopic ultrasonography-guided biliary drainage is an evolving method of biliary drainage in patients with malignant biliary obstruction, particularly in patients whom standard endoscopic retrograde cholangiopancreatography failed.

SPINAL CORD COMPRESSION

Malignant spinal cord compression (MSCC) is defined as compression of the spinal cord and/or cauda equina by an extradural tumor mass. The minimum radiologic evidence for cord compression is indentation of the theca at the level of clinical features. Spinal cord compression

(SCC) occurs in 5–10% of patients with cancer. Epidural tumor is the first manifestation of malignancy in ~10% of patients. The underlying cancer is usually identified during the initial evaluation; lung cancer is the most common cause of MSCC.

Metastatic tumor involves the vertebral column more often than any other part of the bony skeleton. Lung, breast, and prostate cancers are the most frequent offenders. Multiple myeloma also has a high incidence of spine involvement. Lymphomas, melanoma, renal cell cancer, and genitourinary cancers also cause cord compression. The thoracic spine is the most common site (70%), followed by the lumbosacral spine (20%) and the cervical spine (10%). Involvement of multiple sites is most frequent in patients with breast and prostate carcinoma. Cord injury develops when metastases to the vertebral body or pedicle enlarge and compress the underlying dura. Another cause of cord compression is direct extension of a paravertebral lesion through the intervertebral foramen. These cases usually involve a lymphoma, myeloma, or pediatric neoplasm. Parenchymal spinal cord metastasis due to hematogenous spread is rare. Intramedullary metastases can be seen in lung cancer, breast cancer, renal cancer, melanoma, and lymphoma, and are frequently associated with brain metastases and leptomeningeal disease.

Expanding extradural tumors induce injury through several mechanisms. Expanding extradural tumors induce mechanical injury to axons and myelin. Compression compromises blood flow, leading to ischemia and/or infarction.

The most common initial symptom in patients with SCC is localized back pain and tenderness due to involvement of vertebrae by tumor. Pain is usually present for days or months before other neurologic findings appear. It is exacerbated by movement and by coughing or sneezing. It can be differentiated from the pain of disk disease by the fact that it worsens when the patient is supine. Radicular pain is less common than localized back pain and usually develops later. Radicular pain in the cervical or lumbosacral areas may be unilateral or bilateral. Radicular pain from the thoracic roots is often bilateral and is described by patients as a feeling of tight, band-like constriction around the thorax and abdomen. Typical cervical radicular pain radiates down the arm; in the lumbar region, the radiation is down the legs. *Lhermitte's sign*, a tingling or electric sensation down the back and upper and lower limbs upon flexing or extending the neck, may be an early sign of cord compression. Loss of bowel or bladder control may be the presenting symptom but usually occurs late in the course. Occasionally, patients present with ataxia of gait without motor and sensory involvement due to involvement of the spinocerebellar tract.

On physical examination, pain induced by straight leg raising, neck flexion, or vertebral percussion may help to determine the level of cord compression. Patients develop numbness and paresthesias in the extremities or trunk. Loss of sensibility to pinprick is as common as loss of sensibility to vibration or position. The upper limit of the zone of sensory loss is often one or two vertebrae below the site of compression. Motor findings include weakness, spasticity, and abnormal muscle stretching. An extensor plantar reflex reflects significant compression. Deep tendon reflexes may be brisk. Motor and sensory loss usually precedes sphincter disturbance. Patients with autonomic dysfunction may present with decreased anal tonus, decreased perineal sensibility, and a distended bladder. The absence of the anal wink reflex or the bulbocavernosus reflex confirms cord involvement. In doubtful cases, evaluation of postvoiding urinary residual volume can be helpful. A residual volume of >150 mL suggests bladder dysfunction. Autonomic dysfunction is an unfavorable prognostic factor. Patients with progressive neurologic symptoms should have frequent neurologic examinations and rapid therapeutic intervention. Other illnesses that may mimic cord compression include osteoporotic vertebral collapse, disk disease, pyogenic abscess or vertebral tuberculosis, radiation myelopathy, neoplastic leptomeningitis, benign tumors, epidural hematoma, and spinal lipomatosis.

Cauda equina syndrome is characterized by low back pain; diminished sensation over the buttocks, posterior-superior thighs, and perineal area in a saddle distribution; rectal and bladder dysfunction; sexual impotence; absent bulbocavernous, patellar, and Achilles'

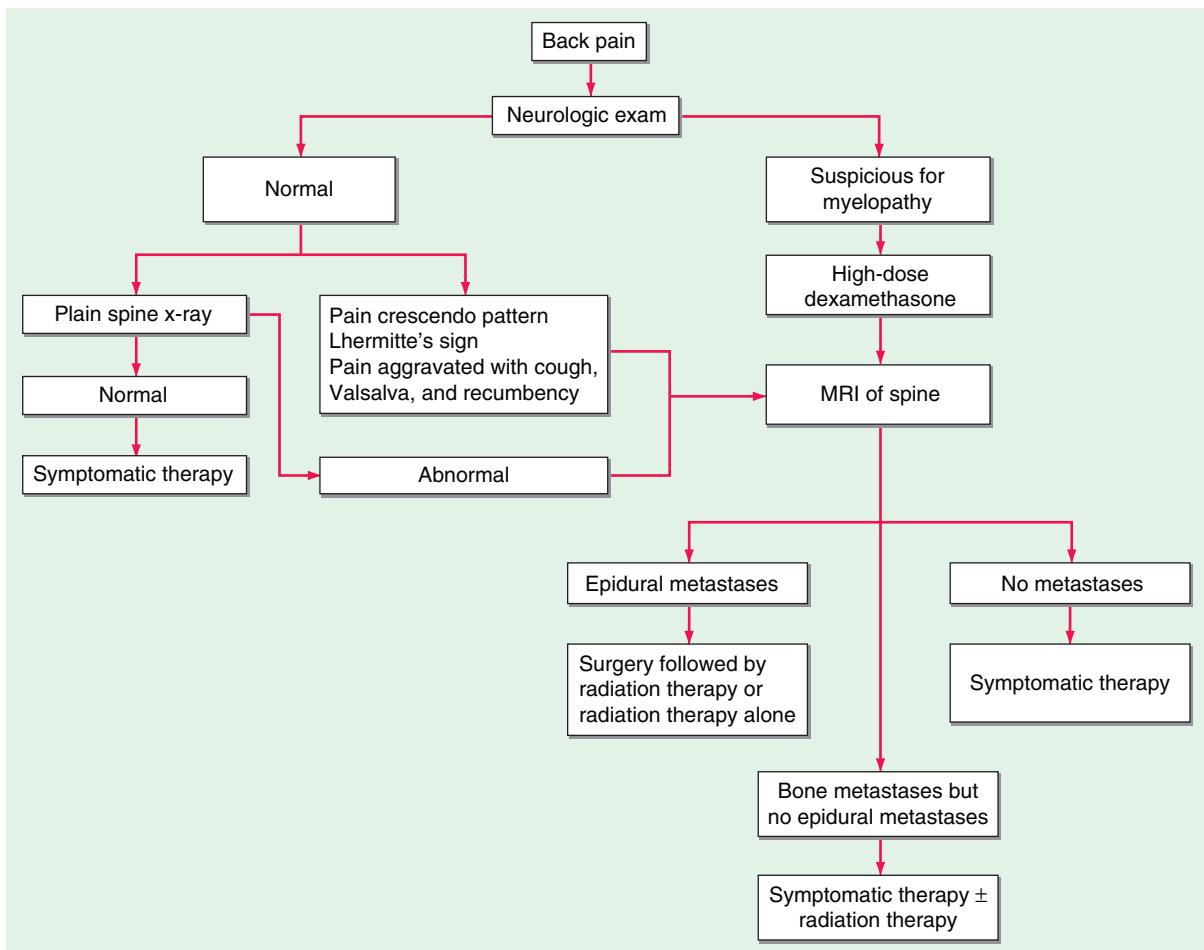


FIGURE 75-2 Management of cancer patients with back pain.

reflexes; and variable amount of lower-extremity weakness. This reflects compression of nerve roots as they form the cauda equina after leaving the spinal cord. The majority of cauda equina tumors are primary tumors of glial or nerve sheath origin; metastases are very rare.

Patients with cancer who develop back pain should be evaluated for SCC as quickly as possible (Fig. 75-2). Treatment is more often successful in patients who are ambulatory and still have sphincter control at the time treatment is initiated. Patients should have a neurologic examination and plain films of the spine. Those whose physical examination suggests cord compression should receive dexamethasone starting immediately and undergo MRI imaging.

Erosion of the pedicles (the “winking owl” sign) is the earliest radiologic finding of vertebral tumor in plain films; however, plain films are insensitive. Other radiographic changes include increased intrapedicular distance, vertebral destruction, lytic or sclerotic lesions, scalloped vertebral bodies, and vertebral body collapse. Vertebral collapse is not a reliable indicator of the presence of tumor; ~20% of cases of vertebral collapse, particularly those in older patients and postmenopausal women, are due not to cancer but to osteoporosis. Also, a normal appearance on plain films of the spine does not exclude the diagnosis of cancer. The role of bone scans in the detection of cord compression is not clear; this method is sensitive but less specific than spinal radiography.

The full-length image of the cord provided by MRI is the imaging procedure of choice. Multiple epidural metastases are noted in 25% of patients with cord compression, and their presence influences treatment plans. On T1-weighted images, good contrast is noted between the cord, cerebrospinal fluid (CSF), and extradural lesions. Owing to its sensitivity in demonstrating the replacement of bone marrow by tumor, MRI can show which parts of a vertebra are involved by tumor. MRI also visualizes intraspinal extradural masses compressing

the cord. T2-weighted images are most useful for the demonstration of intramedullary pathology. Gadolinium-enhanced MRI can help to delineate intramedullary disease. MRI is as good as or better than myelography plus postmyelogram CT scan in detecting metastatic epidural disease with cord compression. Myelography should be reserved for patients who have poor MRIs or who cannot undergo MRI promptly. CT scan in conjunction with myelography enhances the detection of small areas of spinal destruction.

In patients with cord compression and an unknown primary tumor, a simple workup including chest radiography, mammography, measurement of prostate-specific antigen, and abdominal CT usually reveals the underlying malignancy.

TREATMENT

Spinal Cord Compression

The treatment of patients with SCC is aimed at relief of pain and restoration/preservation of neurologic function (Fig. 75-2). Management of MSCC requires a multidisciplinary approach.

Radiation therapy plus glucocorticoids is generally the initial treatment of choice for most patients with SCC. The management decision of SCC involves assessment of neurologic (N), oncologic (O), mechanical (M), and systemic factors (S). NOMS was developed by Memorial Sloan Kettering Cancer Center (MSKCC) researchers to provide an algorithm for management of SCC. The neurologic assessment is based on the degree of epidural SCC, myelopathy, and/or functional radiculopathy. Oncologic assessment involves the radiosensitivity of the tumor type. In patients with radioresistant tumors, stereotactic body radiotherapy (SBRT) is the preferred approach if radiation is appropriate. Safe delivery

of SBRT requires a 2- to 3-mm margin away from the spinal cord. Separation surgery followed by SBRT is necessary in patients with high-grade SCC due to radioresistant tumors. Separation surgery is the circumferential excision of epidural tumor to reconstitute the thecal sac and provide a 2-mm margin for safe delivery of an ablative radiation dose. In patients with mechanical instability or retropulsion of bone fragments into the spinal canal or cord, a surgical approach is the treatment of choice. Systemic factors that need to be considered are the extent of disease and medical comorbidities that determine the patient's ability to tolerate planned therapy. Chemotherapy may have a role in patients with chemosensitive tumors who have had prior radiotherapy to the same region and who are not candidates for surgery. Patients who previously received radiotherapy for MSCC with an in-field tumor progression can be treated with reirradiation with spine stereotactic radiosurgery (SRS) if they are not surgical candidates.

Patients with painful pathologic compression fractures without spinal instability may benefit from percutaneous vertebroplasty or kyphoplasty, the injection of acrylic cement into a collapsed vertebra to stabilize the fracture. Pain palliation is common, and local antitumor effects have been noted. Cement leakage may cause symptoms in ~10% of patients. Bisphosphonates and/or denosumab may be helpful in prevention of SCC in patients with bony involvement.

The histology of the tumor is an important determinant of both recovery and survival. Rapid onset and progression of signs and symptoms are poor prognostic features.

INCREASED INTRACRANIAL PRESSURE

About 25% of patients with cancer die with intracranial metastases. The cancers that most often metastasize to the brain are lung and breast cancers and melanoma. Brain metastases often occur in the presence of systemic disease, and they frequently cause major symptoms, disability, and early death. The initial presentation of brain metastases from a previously unknown primary cancer is common. Lung cancer is most commonly the primary malignancy. CT scans of the chest/abdomen and MRI of the brain as the initial diagnostic studies can identify a biopsy site in most patients.

The signs and symptoms of a metastatic brain tumor are similar to those of other intracranial expanding lesions: headache, nausea, vomiting, behavioral changes, seizures, and focal, progressive neurologic changes. Occasionally the onset is abrupt, resembling a stroke, with the sudden appearance of headache, nausea, vomiting, and neurologic deficits. This picture is usually due to hemorrhage into the metastasis. Melanoma, germ cell tumors, and renal cell cancers have a particularly high incidence of intracranial bleeding. The tumor mass and surrounding edema may cause obstruction of the circulation of CSF, with resulting hydrocephalus. Patients with increased intracranial pressure may have papilledema with visual disturbances and neck stiffness. As the mass enlarges, brain tissue may be displaced through the fixed cranial openings, producing various herniation syndromes.

MRI is superior to CT scan. Gadolinium-enhanced MRI is more sensitive than CT at revealing meningeal involvement and small lesions, particularly in the brainstem or cerebellum. The MRI of the brain shows brain metastases as multiple enhancing lesions of various sizes with surrounding areas of low-density edema.

Intracranial hypertension ("pseudotumor cerebri") secondary to tretinoin therapy for acute promyelocytic leukemia has been reported as another cause of intracranial pressure in the setting of a malignancy.

TREATMENT

Increased Intracranial Pressure

Dexamethasone is the best initial treatment for all symptomatic patients with brain metastases. The current success of immunotherapy approaches for primary and metastatic brain tumors may preclude or limit glucocorticoid use since it may decrease antitumor response. Bevacizumab should be considered in patients who are unable to wean completely off of steroids as well as those who have

symptomatic brain edema and are on immunotherapy. Patients with multiple lesions should usually receive whole-brain radiation. Patients with a single brain metastasis and with controlled extracranial disease may be treated with surgical excision followed by whole-brain radiation therapy, especially if they are aged <60 years. Radioresistant tumors should be resected if possible. Stereotactic radiosurgery (SRS) is recommended in patients with a limited number of brain metastases (one to four) who have stable, systemic disease or reasonable systemic treatment options and in patients who have a small number of metastatic lesions in whom whole-brain radiation therapy has failed. With a gamma knife or linear accelerator, multiple small, well-collimated beams of ionizing radiation destroy lesions seen on MRI. Some patients with increased intracranial pressure associated with hydrocephalus may benefit from shunt placement. If neurologic deterioration is not reversed with medical therapy, ventriculotomy to remove CSF or craniotomy to remove tumors or hematomas may be necessary.

Targeted agents and checkpoint inhibitors have significant activity in brain metastases from non-small-cell lung cancer, breast cancer, renal cancer, and melanoma.

NEOPLASTIC MENINGITIS

Tumor involving the leptomeninges is a complication of both primary central nervous system (CNS) tumors and tumors that metastasize to the CNS. The incidence is estimated at 3–8% of patients with cancer. Melanoma, breast and lung cancer, lymphoma (including AIDS-associated), and acute leukemia are the most common causes. The lobular or triple-negative subtypes of breast cancer, as well as tumors with expression of the mutant epidermal growth factor receptor (EGFR) or the anaplastic lymphoma kinase (ALK) rearrangement in non-small-cell lung cancer (NSCLC), are more likely to have CNS involvement including neoplastic meningitis and brain metastases. Synchronous intraparenchymal brain metastases are evident in 11–31% of patients with neoplastic meningitis. Leptomeningeal seeding is frequent in patients undergoing resection of brain metastases or receiving stereotactic radiotherapy for brain metastases.

Patients typically present with multifocal neurologic signs and symptoms, including headache, gait abnormality, mental changes, nausea, vomiting, seizures, back or radicular pain, and limb weakness. Signs include cranial nerve palsies, extremity weakness, paresthesia, and decreased deep tendon reflexes.

Diagnosis is made by demonstrating malignant cells in the CSF; however, up to 40% of patients may have false-negative CSF cytology. An elevated CSF protein level is nearly always present (except in HTLV-1-associated adult T-cell leukemia). Patients with neurologic signs and symptoms consistent with neoplastic meningitis who have a negative CSF cytology should have the spinal tap repeated at least one more time for cytologic examination. MRI findings suggestive of neoplastic meningitis include leptomeningeal, subependymal, dural, or cranial nerve enhancement; superficial cerebral lesions; intradural nodules; and communicating hydrocephalus. Spinal cord imaging by MRI is a necessary component of the evaluation of nonleukemia neoplastic meningitis because ~20% of patients have cord abnormalities, including intradural enhancing nodules that are diagnostic for leptomeningeal involvement. Cauda equina lesions are common, but lesions may be seen anywhere in the spinal canal. The value of MRI for the diagnosis of leptomeningeal disease is limited in patients with hematopoietic malignancy. Radiolabeled CSF flow studies are abnormal in up to 70% of patients with neoplastic meningitis; ventricular outlet obstruction, abnormal flow in the spinal canal, or impaired flow over the cerebral convexities may affect distribution of intrathecal chemotherapy, resulting in decreased efficacy or increased toxicity. Radiation therapy may correct CSF flow abnormalities before use of intrathecal chemotherapy. Neoplastic meningitis can also lead to intracranial hypertension and hydrocephalus. Placement of a ventriculoperitoneal shunt may effectively palliate symptoms in these patients.

The development of neoplastic meningitis usually occurs in the setting of uncontrolled cancer outside the CNS; thus, prognosis is poor

(median survival 10–12 weeks). However, treatment of the neoplastic meningitis may successfully alleviate symptoms and control the CNS spread.

TREATMENT

Neoplastic Meningitis

Chemotherapy provided by either intrathecal injection or systemic routes is used to control leptomeningeal disease throughout the entire neuroaxis. Intrathecal chemotherapy, usually methotrexate, cytarabine, or thiotapec, is delivered by lumbar puncture or by an intraventricular reservoir (Ommaya). Among solid tumors, breast cancer responds best to therapy. Focal radiotherapy may have a role in bulky disease and in symptomatic or obstructive lesions. Targeted therapy such as systemically administered EGFR tyrosine kinase inhibitors (TKIs) in non-small-cell lung cancer may lead to improvement in some patients with leptomeningeal spread. Patients with neoplastic meningitis from either acute leukemia or lymphoma may be cured of their CNS disease if the systemic disease can be eliminated.

SEIZURES

Seizures occurring in a patient with cancer can be caused by the tumor itself, by metabolic disturbances, by radiation injury, by cerebral infarctions, by chemotherapy-related encephalopathies, or by CNS infections. Metastatic disease to the CNS is the most common cause of seizures in patients with cancer. However, seizures occur more frequently in primary brain tumors than in metastatic brain lesions. Seizures are a presenting symptom of CNS metastasis in 6–29% of cases. Approximately 10% of patients with CNS metastasis eventually develop seizures. Tumors that affect the frontal, temporal, and parietal lobes are more commonly associated with seizures than are occipital lesions. Both early and late seizures are uncommon in patients with posterior fossa and sellar lesions. Seizures are common in patients with CNS metastases from melanoma and low-grade primary brain tumors. Very rarely, cytotoxic drugs such as etoposide, busulfan, ifosfamide, and chlorambucil cause seizures. Treatment with bispecific antibodies and chimeric antigen receptor (CAR) T cells may also cause CNS toxicity including seizures and encephalopathy. Another cause of seizures related to drug therapy is reversible posterior leukoencephalopathy syndrome (RPLS). Chemotherapy, targeted therapy, and immunotherapies have been associated with the development of RPLS. RPLS occurs in patients undergoing allogeneic bone marrow or solid-organ transplantation. RPLS is characterized by headache, altered consciousness, generalized seizures, visual disturbances, hypertension, and symmetric posterior cerebral white matter vasogenic edema on CT/MRI. Seizures may begin focally but are typically generalized.

TREATMENT

Seizures

Patients in whom seizures due to CNS metastases have been demonstrated should receive anticonvulsive treatment with phenytoin or levetiracetam. If this is not effective, valproic acid can be added. Prophylactic anticonvulsant therapy is not recommended. In postcraniotomy patients, prophylactic antiepileptic drugs should be withdrawn during the first week after surgery. Most antiseizure medications including phenytoin induce cytochrome P450 (CYP450), which alters the metabolism of many antitumor agents, including irinotecan, taxanes, and etoposide as well as molecular targeted agents, including imatinib, gefitinib, erlotinib, tipifarnib, sorafenib, sunitinib, temsirolimus, everolimus, and vemurafenib. Levetiracetam and topiramate are anticonvulsant agents not metabolized by the hepatic CYP450 system and do not alter the metabolism of antitumor agents. They have become the preferred drugs. Surgical resection and other antitumor treatments such as radiotherapy and chemotherapy may improve seizure control.

PULMONARY AND INTRACEREBRAL LEUKOSTASIS

Hyperleukocytosis and the leukostasis syndrome associated with it are potentially fatal complications of acute leukemia (particularly myeloid leukemia) that can occur when the peripheral blast cell count is $>100,000/\text{mL}$. The frequency of hyperleukocytosis is 5–13% in acute myeloid leukemia (AML) and 10–30% in acute lymphoid leukemia; however, leukostasis is rare in lymphoid leukemia. At such high blast cell counts, blood viscosity is increased, blood flow is slowed by aggregates of tumor cells, and the primitive myeloid leukemic cells are capable of invading through the endothelium and causing hemorrhage. Brain and lung are most commonly affected. Patients with brain leukostasis may experience stupor, headache, dizziness, tinnitus, visual disturbances, ataxia, confusion, coma, or sudden death. On examination, papilledema, retinal vein distension, retinal hemorrhages, and focal deficit may be present. Pulmonary leukostasis may present as respiratory distress and hypoxemia and progress to respiratory failure. Chest radiographs may be normal but usually show interstitial or alveolar infiltrates. Hyperleukocytosis rarely may cause acute leg ischemia, renal vein thrombosis, myocardial ischemia, bowel infarction, and priapism. Arterial blood gas results should be interpreted cautiously. Rapid consumption of plasma oxygen by the markedly increased number of white blood cells can cause spuriously low arterial oxygen tension. Pulse oximetry is the most accurate way of assessing oxygenation in patients with hyperleukocytosis. Hydroxyurea can rapidly reduce a high blast cell count while the diagnostic workup is in progress. After the diagnosis is established, the patient should start quickly with effective induction chemotherapy. Leukapheresis should be used in patients with symptoms of hyperleukocytosis. Patients with hyperleukocytosis are also at the risk for disseminated intravascular coagulation and tumor lysis syndrome. The clinician should monitor the patient for these complications and take preventive and therapeutic actions during induction therapy. Intravascular volume depletion and unnecessary blood transfusions may increase blood viscosity and worsen the leukostasis syndrome. Leukostasis is very rarely a feature of the high white cell counts associated with chronic lymphoid or chronic myeloid leukemia.

When acute promyelocytic leukemia is treated with differentiating agents like tretinoin and arsenic trioxide, cerebral or pulmonary leukostasis may occur as tumor cells differentiate into mature neutrophils. This complication can be largely avoided by using cytotoxic chemotherapy together with the differentiating agents.

HEMOPTYSIS

Hemoptysis may be caused by nonmalignant conditions, but lung cancer accounts for a large proportion of cases. Up to 20% of patients with lung cancer have hemoptysis some time in their course. Endobronchial metastases from carcinoid tumors, breast cancer, colon cancer, kidney cancer, and melanoma may also cause hemoptysis. The volume of bleeding is often difficult to gauge. Massive hemoptysis is defined as $>200\text{--}600 \text{ mL}$ of blood produced in 24 h. However, any hemoptysis should be considered massive if it threatens life. When respiratory difficulty occurs, hemoptysis should be treated emergently. The first priorities are to maintain the airway, optimize oxygenation, and stabilize the hemodynamic status. If the bleeding side is known, the patient should be placed in a lateral decubitus position, with the bleeding side down to prevent aspiration into the unaffected lung, and given supplemental oxygen. If large-volume bleeding continues or the airway is compromised, the patient should be intubated and undergo emergency bronchoscopy. If the site of bleeding is detected, either the patient undergoes a definitive surgical procedure or the lesion is treated with a neodymium:yttrium-aluminum-garnet (Nd:YAG) laser, argon plasma coagulation, or electrocautery. In stable patients, multidetector CT angiography delineates bronchial and nonbronchial systemic arteries and identifies the source of bleeding and underlying pathology with high sensitivity. Massive hemoptysis usually originates from the high-pressure bronchial circulation. Bronchial artery embolization is considered a first-line definitive procedure for managing hemoptysis. Bronchial artery embolization may control brisk bleeding in 75–90% of

patients, permitting the definitive surgical procedure to be done more safely if it is appropriate.

Embolization without definitive surgery is associated with rebleeding in 20–50% of patients. Recurrent hemoptysis usually responds to a second embolization procedure. A postembolization syndrome characterized by pleuritic pain, fever, dysphagia, and leukocytosis may occur; it lasts 5–7 days and resolves with symptomatic treatment. Bronchial or esophageal wall necrosis, myocardial infarction, and spinal cord infarction are rare complications. Surgery, as a salvage strategy, is indicated after failure of embolization and is associated with better survival when performed in a nonurgent setting.

Pulmonary hemorrhage with or without hemoptysis in hematologic malignancies is often associated with fungal infections, particularly *Aspergillus* spp. After granulocytopenia resolves, the lung infiltrates in aspergillosis may cavitate and cause massive hemoptysis. Thrombocytopenia and coagulation defects should be corrected, if possible. Surgical evaluation is recommended in patients with aspergillosis-related cavitary lesions.

Bevacizumab, an antibody to vascular endothelial growth factor (VEGF) that inhibits angiogenesis, has been associated with life-threatening hemoptysis in patients with non-small-cell lung cancer, particularly of squamous cell histology. Non-small-cell lung cancer patients with cavitary lesions or previous hemoptysis (≥ 2.5 mL) within the past 3 months have higher risk for pulmonary hemorrhage.

AIRWAY OBSTRUCTION

Airway obstruction refers to a blockage at the level of the mainstem bronchi or above. It may result either from intraluminal tumor growth or from extrinsic compression of the airway. The most common cause of malignant upper airway obstruction is invasion from an adjacent primary tumor, most commonly lung cancer, followed by esophageal, thyroid, and mediastinal malignancies including lymphomas. Extrathoracic primary tumors such as renal, colon, or breast cancer can cause airway obstruction through endobronchial and/or mediastinal lymph node metastases. Patients may present with dyspnea, hemoptysis, stridor, wheezing, intractable cough, postobstructive pneumonia, or hoarseness. Chest radiographs usually demonstrate obstructing lesions. CT scans reveal the extent of tumor. Cool, humidified oxygen, glucocorticoids, and ventilation with a mixture of helium and oxygen (Heliox) may provide temporary relief. If the obstruction is proximal to the larynx, a tracheostomy may be lifesaving. For more distal obstructions, particularly intrinsic lesions incompletely obstructing the airway, bronchoscopy with mechanical debulking and dilation or ablational treatments including laser treatment, photodynamic therapy, argon plasma coagulation, electrocautery, or stenting can produce immediate relief in most patients (Fig. 75-3). However, radiation therapy (either external-beam irradiation or brachytherapy) given together with glucocorticoids may also open the airway. Symptomatic extrinsic compression may be palliated by stenting. Patients with primary airway tumors such as squamous cell carcinoma, carcinoid tumor, adenocystic carcinoma, or non-small-cell lung cancer, if resectable, should have surgery.

METABOLIC EMERGENCIES

HYPERCALCEMIA

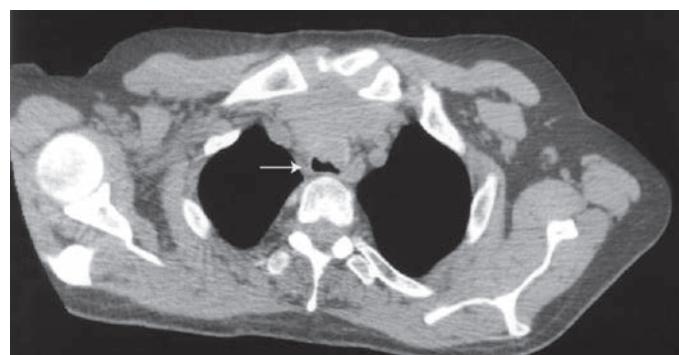
Hypercalcemia is the most common paraneoplastic syndrome. Its pathogenesis and management are discussed fully in Chaps. 93 and 410.

SYNDROME OF INAPPROPRIATE SECRETION OF ANTIIDIURETIC HORMONE

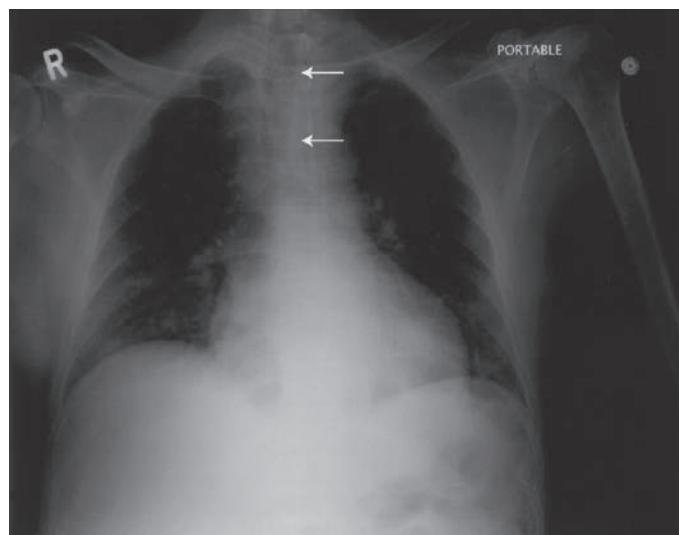
Hyponatremia is a common electrolyte abnormality in cancer patients, and syndrome of inappropriate secretion of antidiuretic hormone (SIADH) is the most common cause among patients with cancer. SIADH is discussed fully in Chaps. 93 and 381.

LACTIC ACIDOSIS

Lactic acidosis is a rare and potentially fatal metabolic complication of cancer. Lactic acidosis associated with sepsis and circulatory failure is a common preterminal event in many malignancies. Lactic acidosis in the



A



B

FIGURE 75-3 Airway obstruction. **A.** Computed tomography scan of a 62-year-old man with tracheal obstruction caused by renal carcinoma showing paratracheal mass with tracheal invasion/obstruction (arrow). **B.** Chest x-ray of same patient after stent (arrows) placement.

absence of hypoxemia may occur in patients with leukemia, lymphoma, or solid tumors. In some cases, hypoglycemia also is present. Extensive involvement of the liver by tumor is often present. In most cases, decreased metabolism and increased production by the tumor both contribute to lactate accumulation. Tumor cell overexpression of certain glycolytic enzymes and mitochondrial dysfunction can contribute to its increased lactate production. HIV-infected patients have an increased risk of aggressive lymphoma; lactic acidosis that occurs in such patients may be related either to the rapid growth of the tumor or from toxicity of nucleoside reverse transcriptase inhibitors. Symptoms of lactic acidosis include tachypnea, tachycardia, change of mental status, and hepatomegaly. The serum level of lactic acid may reach 10–20 mmol/L (90–180 mg/dL). Treatment is aimed at the underlying disease. *The danger from lactic acidosis is from the acidosis, not the lactate.* Sodium bicarbonate should be added if acidosis is very severe or if hydrogen ion production is very rapid and uncontrolled. Other treatment options include renal replacement therapy, such as hemodialysis, and thiamine replacement. The prognosis is poor regardless of the treatment offered.

HYPOGLYCEMIA

Persistent hypoglycemia is occasionally associated with tumors other than pancreatic islet cell tumors. Usually these tumors are large; tumors of mesenchymal origin, hepatomas, or adrenocortical tumors may cause hypoglycemia. Mesenchymal tumors are usually located in the retroperitoneum or thorax. Obtundation, confusion, and behavioral aberrations occur in the postabsorptive period and may precede the diagnosis of the tumor. These tumors often secrete incompletely processed insulin-like growth factor II (IGF-II), a hormone capable

of activating insulin receptors and causing hypoglycemia. Tumors secreting incompletely processed big IGF-II are characterized by an increased IGF-II to IGF-I ratio, suppressed insulin and C-peptide level, and inappropriately low growth hormone and β -hydroxybutyrate concentrations. Rarely, hypoglycemia is due to insulin secretion by a non-islet cell carcinoma. The development of hepatic dysfunction from liver metastases and increased glucose consumption by the tumor can contribute to hypoglycemia. If the tumor cannot be resected, hypoglycemia symptoms may be relieved by the administration of glucose, glucocorticoids, recombinant growth hormone, or glucagon.

Hypoglycemia can be artificial; hyperleukocytosis from leukemia, myeloproliferative diseases, leukemoid reactions, or colony-stimulating factor treatment can increase glucose consumption in the test tube after blood is drawn, leading to pseudohypoglycemia.

■ ADRENAL INSUFFICIENCY

In patients with cancer, adrenal insufficiency may go unrecognized because the symptoms, such as nausea, vomiting, anorexia, and orthostatic hypotension, are nonspecific and may be mistakenly attributed to progressive cancer or to therapy. Primary adrenal insufficiency may develop owing to replacement of both glands by metastases (lung, breast, colon, or kidney cancer; lymphoma), to removal of both glands, or to hemorrhagic necrosis in association with sepsis or anticoagulation. Impaired adrenal steroid synthesis occurs in patients being treated for cancer with mitotane, ketoconazole, or aminoglutethimide or undergoing rapid reduction in glucocorticoid therapy. Megestrol acetate, used to manage cancer and HIV-related cachexia, may suppress plasma levels of cortisol and adrenocorticotrophic hormone (ACTH). Patients taking megestrol may develop adrenal insufficiency, and even those whose adrenal dysfunction is not symptomatic may have inadequate adrenal reserve if they become seriously ill. Paradoxically, some patients may develop Cushing's syndrome and/or hyperglycemia because of the glucocorticoid-like activity of megestrol acetate. Ipilimumab, an anti-CTLA-4 antibody used for treatment of malignant melanoma, may cause autoimmunity including autoimmune-like enterocolitis, hypophysitis (leading to secondary adrenal insufficiency), hepatitis, and, rarely, primary adrenal insufficiency. Autoimmune hypophysitis may present with headache, visual field defects, and pituitary hormone deficiencies manifesting as hypopituitarism, adrenal insufficiency (including adrenal crisis), or hypothyroidism. Ipilimumab-associated hypophysitis symptoms occur at an average of 6–12 weeks after initiation of therapy. An MRI usually shows homogenous enhancement of pituitary gland. Early glucocorticoid treatment and hormone replacement are the initial treatment. The role of high-dose glucocorticoids in the treatment of hypophysitis is not clear. High-dose glucocorticoids may not improve the frequency of pituitary function recovery. Autoimmune adrenalitis can also be observed with anti-CTLA-4 antibody. Pituitary dysfunction is usually permanent, requiring long-term hormone replacement therapy. Other checkpoint inhibitors, such as monoclonal antibodies targeting programmed cell death-1 (PD-1), an inhibitory receptor expressed by T cells or one of its ligands (PD-L1), may cause hypophysitis infrequently (~1%). Autoimmune adrenalitis is more frequent with use of PD-1/PD-L1 than with CTLA-4 inhibitors, but incidence is low. Cranial irradiation for childhood brain tumors may affect the hypothalamus-pituitary-adrenal axis, resulting in secondary adrenal insufficiency. Rarely, metastatic replacement causes primary adrenal insufficiency as the first manifestation of an occult malignancy. Metastasis to the pituitary or hypothalamus is found at autopsy in up to 5% of patients with cancer, but associated secondary adrenal insufficiency is rare.

Acute adrenal insufficiency is potentially lethal. Treatment of suspected adrenal crisis is initiated after the sampling of serum cortisol and ACTH levels ([Chap. 386](#)).

TREATMENT-RELATED EMERGENCIES

■ TUMOR LYYSIS SYNDROME

Tumor lysis syndrome (TLS) is characterized by hyperuricemia, hyperkalemia, hyperphosphatemia, and hypocalcemia and is caused by the

destruction of a large number of rapidly proliferating neoplastic cells. Acidosis may also develop. Acute renal failure occurs frequently.

TLS is most often associated with the treatment of Burkitt's lymphoma, acute lymphoblastic leukemia, and other rapidly proliferating lymphomas, but it also may be seen with chronic leukemias and, rarely, with solid tumors. This syndrome has been seen in patients with chronic lymphocytic leukemia after treatment with nucleosides like fludarabine and is increased in frequency in lymphoid neoplasms treated with venetoclax, a bcl-2 antagonist. TLS has been observed with administration of glucocorticoids, hormonal agents such as letrozole and tamoxifen, and monoclonal antibodies such as rituximab, obinutuzumab, ofatumumab, and gemtuzumab. TLS usually occurs during or shortly (1–5 days) after chemotherapy. Rarely, spontaneous necrosis of malignancies causes TLS.

Hyperuricemia may be present at the time of chemotherapy. Effective treatment kills malignant cells and leads to increased serum uric acid levels from the turnover of nucleic acids. Owing to the acidic local environment, uric acid can precipitate in the tubules, medulla, and collecting ducts of the kidney, leading to renal failure. Lactic acidosis and dehydration may contribute to the precipitation of uric acid in the renal tubules. The finding of uric acid crystals in the urine is strong evidence for uric acid nephropathy. The ratio of urinary uric acid to urinary creatinine is >1 in patients with acute hyperuricemic nephropathy and <1 in patients with renal failure due to other causes. Other events may lead to renal failure in TLS. Calcium phosphate also precipitates in the interstitium and renal microvasculature, leading to nephrocalcinosis. Both types of crystals are toxic to the tubular epithelium, inducing local active inflammatory and pro-oxidative responses. Soluble uric acid may induce hemodynamic changes, with decreased renal blood flow due to vasoconstriction and impaired autoregulation (crystal-independent pathway).

Hyperphosphatemia, which can be caused by the release of intracellular phosphate pools by tumor lysis, produces a reciprocal depression in serum calcium, which causes severe neuromuscular irritability and tetany. Deposition of calcium phosphate in the kidney and hyperphosphatemia may cause renal failure. Potassium is the principal intracellular cation, and massive destruction of malignant cells may lead to hyperkalemia. Hyperkalemia in patients with renal failure may rapidly become life threatening by causing ventricular arrhythmias and sudden death.

The likelihood that TLS will occur in patients with Burkitt's lymphoma is related to the tumor burden and renal function. Hyperuricemia and high serum levels of lactate dehydrogenase (LDH >1500 U/L), both of which correlate with total tumor burden, also correlate with the risk of TLS. In patients at risk for TLS, pretreatment evaluations should include a complete blood count, serum chemistry evaluation, and urine analysis. High leukocyte and platelet counts may artificially elevate potassium levels ("pseudohyperkalemia") due to lysis of these cells after the blood is drawn. In these cases, plasma potassium instead of serum potassium should be followed. In pseudohyperkalemia, no electrocardiographic abnormalities are present. In patients with abnormal baseline renal function, the kidneys and retroperitoneal area should be evaluated by sonography and/or CT to rule out obstructive uropathy. Urine output should be watched closely.

TREATMENT

Tumor Lysis Syndrome

Recognition of risk and prevention are the most important steps in the management of this syndrome ([Fig. 75-4](#)). The standard preventive approach consists of allopurinol and aggressive hydration. Urinary alkalization with sodium bicarbonate is no longer recommended. It increases uric acid solubility, but a high pH decreases the solubility of xanthine, hypoxanthine, and calcium phosphate, potentially increasing the likelihood of intratubular crystallization. Intravenous allopurinol may be given in patients who cannot tolerate oral therapy. Febuxostat, a potent nonpurine selective xanthine oxidase inhibitor, is indicated for treatment of hyperuricemia. It

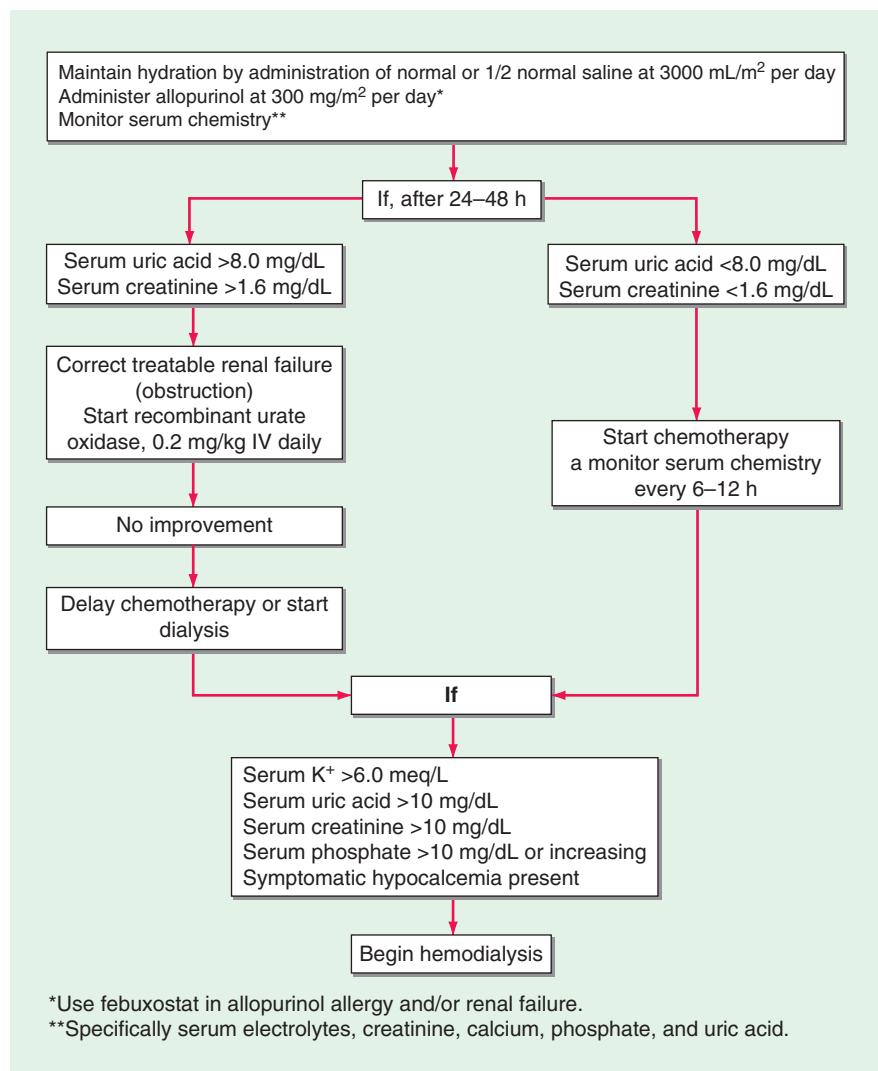


FIGURE 75-4 Management of patients at high risk for the tumor lysis syndrome.

results in fewer hypersensitivity reactions than allopurinol. Febuxostat does not require dosage adjustment in patients with mild to moderate renal impairment. Febuxostat achieved significantly superior serum uric acid control in comparison to allopurinol in patients with hematologic malignancies at intermediate to high TLS risk. In some cases, uric acid levels cannot be lowered sufficiently with the standard preventive approach. Rasburicase (recombinant urate oxidase) can be effective in these instances, particularly when renal failure is present. Urate oxidase is missing from primates and catalyzes the conversion of poorly soluble uric acid to readily soluble allantoic acid. Rasburicase acts rapidly, decreasing uric acid levels within hours; however, it may cause hypersensitivity reactions such as bronchospasm, hypoxemia, and hypotension. Rasburicase should also be administered to high-risk patients for TLS prophylaxis. Rasburicase is contraindicated in patients with glucose-6-phosphate dehydrogenase deficiency who are unable to break down hydrogen peroxide, an end product of the urate oxidase reaction. Rasburicase is known to cause ex vivo enzymatic degradation of uric acid in test tube at room temperature. This leads to spuriously low uric acid levels during laboratory monitoring of the patient with TLS. Samples must be cooled immediately to deactivate the urate oxidase. Despite aggressive prophylaxis, TLS and/or oliguric or anuric renal failure may occur. Renal replacement therapy is often necessary and should be considered early in the course. Hemodialysis is preferred. Hemofiltration offers a gradual, continuous method of removing cellular by-products and fluid.

HUMAN ANTIBODY INFUSION REACTIONS

The initial infusion of human or humanized antibodies (e.g., rituximab, gemtuzumab, trastuzumab, alemtuzumab, panitumumab, brentuximab vedotin, blinatumomab) is associated with fever, chills, nausea, asthenia, and headache in up to half of treated patients. Bronchospasm and hypotension occur in 1% of patients. Severe manifestations including pulmonary infiltrates, acute respiratory distress syndrome (ARDS), and cardiogenic shock occur rarely. Laboratory manifestations include elevated hepatic aminotransferase levels, thrombocytopenia, and prolongation of prothrombin time. The pathogenesis is thought to be activation of immune effector processes (cells and complement) and release of inflammatory cytokines, such as tumor necrosis factor α , interferon γ , interleukin (IL) 6, and IL-10 (cytokine release syndrome [CRS]). Although its origins are not completely understood, CRS is believed to be due to activation of a variety of cell types including monocytes/macrophages and T and B lymphocytes. Hemophagocytic lymphohistiocytosis (HLH)/macrophage activation syndrome (MAS) can develop as part of CRS and usually is a manifestation of severe CRS.

Severe CRS may require intensive support for ARDS and resistant hypotension. Emerging clinical experience at several institutions has concluded that tocilizumab is an effective treatment for severe or life-threatening CRS. Tocilizumab prevents IL-6 binding to both cell-associated and soluble IL-6 receptors and therefore inhibits both classical and trans-IL-6 signaling. Other cytokine-directed therapies, such as siltuximab, a chimeric anti-IL-6 monoclonal antibody, and anakinra, an IL-1 receptor antagonist, have been used.

Adoptive transfer of CAR-engineered T cells is a promising therapy for cancers. The most common acute toxicity of CAR T cells is CRS. CAR T cell-associated CRS may be associated with cardiac dysfunction and neurotoxicity. The management includes supportive care and tocilizumab.

Severe reactions from rituximab have occurred with high numbers ($>50 \times 10^9$ lymphocytes) of circulating cells bearing the target antigen (CD20) and have been associated with a rapid fall in circulating tumor cells, mild electrolyte evidence of TLS, and, very rarely, death. In addition, increased liver enzymes, D-dimer, and LDH and prolongation of the prothrombin time may occur. Diphenhydramine, hydrocortisone, and acetaminophen can often prevent or suppress the infusion-related symptoms. If they occur, the infusion is stopped and restarted at half the initial infusion rate after the symptoms have abated.

■ HEMOLYTIC-UREMIC SYNDROME

Hemolytic-uremic syndrome (HUS) and, less commonly, thrombotic thrombocytopenic purpura (TTP) (Chap. 317) may rarely occur after treatment with antineoplastic drugs, including mitomycin, gemcitabine, cisplatin, bleomycin, and proteasome inhibitors, and with VEGF inhibitors. Mitomycin and gemcitabine are the most common offenders. Unlike mitomycin, there is no clear-cut relationship between the cumulative dose of gemcitabine and risk of HUS. It occurs most often in patients with gastric, lung, colorectal, pancreatic, and breast carcinoma. In one series, 35% of patients were without evident cancer at the time this syndrome appeared. Secondary HUS/TTP has also been reported as a rare but sometimes fatal complication of bone marrow transplantation.

HUS usually has its onset 4–8 weeks after the last dose of chemotherapy, but it is not rare to detect it several months later. HUS is characterized by microangiopathic hemolytic anemia, thrombocytopenia, and renal failure. Dyspnea, weakness, fatigue, oliguria, and purpura are also common initial symptoms and findings. Systemic hypertension and pulmonary edema frequently occur. Severe hypertension, pulmonary edema, and rapid worsening of hemolysis and renal function may occur after a blood or blood product transfusion. Cardiac findings include atrial arrhythmias, pericardial friction rub, and pericardial effusion. Raynaud's phenomenon is part of the syndrome in patients treated with bleomycin.

Laboratory findings include severe to moderate anemia associated with red blood cell fragmentation and numerous schistocytes on peripheral smear. Reticulocytosis, decreased plasma haptoglobin, and an LDH level document hemolysis. The serum bilirubin level is usually normal or slightly elevated. The Coombs test is negative. The white cell count is usually normal, and thrombocytopenia ($<100,000/\mu\text{L}$) is almost always present. Most patients have a normal coagulation profile, although some have mild elevations in thrombin time and in levels of fibrin degradation products. The serum creatinine level is elevated at presentation and shows a pattern of subacute worsening within weeks of the initial azotemia. The urinalysis reveals hematuria, proteinuria, and granular or hyaline casts, and circulating immune complexes may be present.

The basic pathologic lesion appears to be deposition of fibrin in the walls of capillaries and arterioles, and these deposits are similar to those seen in HUS due to other causes. These microvascular abnormalities involve mainly the kidneys and rarely occur in other organs. The pathogenesis of cancer treatment-related HUS is not completely understood, but probably the most important factor is endothelial damage. Primary forms of HUS/TTP are related to a decrease in processing of von Willebrand factor by a protease called ADAMTS13.

The case-fatality rate is high; most patients die within a few months. Optimal treatment for chemotherapy-induced HUS is debated. Immunocomplex removal through plasmapheresis, plasma exchange, immunoadsorption, or exchange transfusion, antiplatelet and anticoagulant therapies, and immunosuppression have all been employed with varying degrees of success.

The outcome with plasma exchange is generally poor, as in many other cases of secondary TTP. Rituximab is successfully used in patients with chemotherapy-induced HUS as well as in ADAMTS13-deficient

TTP. Eculizumab, a complement inhibitor, is now approved by the U.S. Food and Drug Administration (FDA) and considered first line for the treatment of atypical HUS. Vaccination against *Neisseria meningitidis* is mandatory before eculizumab is administered.

■ NEUTROPENIA AND INFECTION

These remain the most common serious complications of cancer therapy. They are covered in detail in Chap. 74.

■ PULMONARY INFILTRATES

Patients with cancer may present with dyspnea associated with diffuse interstitial infiltrates on chest radiographs. Such infiltrates may be due to progression of the underlying malignancy, treatment-related toxicities, infection, and/or unrelated diseases. The cause may be multifactorial; however, most commonly, they occur as a consequence of treatment. Infiltration of the lung by malignancy has been described in patients with leukemia, lymphoma, and breast and other solid cancers. Pulmonary lymphatics may be involved diffusely by neoplasm (pulmonary lymphangitic carcinomatosis), resulting in a diffuse increase in interstitial markings on chest radiographs. The patient is often mildly dyspneic at the onset, but pulmonary failure develops over a period of weeks. In some patients, dyspnea precedes changes on the chest radiographs and is accompanied by a nonproductive cough. This syndrome is characteristic of solid tumors. In patients with leukemia, diffuse microscopic neoplastic peribronchial and peribronchiolar infiltration is frequent but may be asymptomatic. However, some patients present with diffuse interstitial infiltrates, an alveolar capillary block syndrome, and respiratory distress. Thickening of bronchovascular bundles and prominence of peripheral arteries are CT findings suggestive of leukemic infiltration. In these situations, glucocorticoids can provide symptomatic relief, but specific chemotherapy should always be started promptly.

Several cytotoxic agents, such as bleomycin, methotrexate, busulfan, nitrosoureas, gemcitabine, mitomycin, vinorelbine, docetaxel, paclitaxel, fludarabine, pentostatin, and ifosfamide, may cause pulmonary damage. The most frequent presentations are interstitial pneumonitis, alveolitis, and pulmonary fibrosis. Some cytotoxic agents, including methotrexate and procarbazine, may cause an acute hypersensitivity reaction. Cytosine arabinoside has been associated with noncardiogenic pulmonary edema. Administration of multiple cytotoxic drugs, as well as radiotherapy and preexisting lung disease, may potentiate the pulmonary toxicity. Supplemental oxygen may potentiate the effects of drugs and radiation injury. Patients should always be managed with the lowest FiO_2 that is sufficient to maintain hemoglobin saturation.

The onset of symptoms may be insidious, with symptoms including dyspnea, nonproductive cough, and tachycardia. Patients may have bibasilar crepitant rales, end-inspiratory crackles, fever, and cyanosis. The chest radiograph generally shows an interstitial and sometimes an intraalveolar pattern that is strongest at the lung bases and may be symmetric. A small effusion may occur. Hypoxemia with decreased carbon monoxide diffusing capacity is always present. Glucocorticoids may be helpful in patients in whom pulmonary toxicity is related to radiation therapy or to chemotherapy. Treatment is otherwise supportive.

Molecular targeted agents, imatinib, erlotinib, and gefitinib are potent inhibitors of tyrosine kinases. These drugs may cause interstitial lung disease (ILD). In the case of gefitinib, preexisting fibrosis, poor performance status, and prior thoracic irradiation are independent risk factors; this complication has a high fatality rate. In Japan, incidence of ILD associated with gefitinib was ~4.5% compared to 0.5% in the United States. Temsirolimus and everolimus, both esters of rapamycin (sirolimus), are agents that block the effects of mammalian target of rapamycin (mTOR), an enzyme that has an important role in regulating the synthesis of proteins that control cell division. These agents may cause ground-glass opacities (GGO) in the lung with or without diffuse interstitial disease and lung parenchymal consolidation. Patients may be asymptomatic with only radiologic findings or may be symptomatic. Symptoms include cough, dyspnea, and/or hypoxemia, and sometimes patients present with systemic symptoms such as fever and fatigue. The incidence of everolimus-induced ILD also appears to be higher in

576 Japanese patients. Treatment includes dose reduction or withdrawal and, in some cases, the addition of glucocorticoids.

The FDA-approved immune checkpoint inhibitors (ICI) of the PD-1 and PD-L1 pathway, including nivolumab, pembrolizumab, durvalumab, avelumab, atezolizumab, and cemiplimab, enhance antitumor activity by blocking negative regulators of T-cell function. Immune-mediated pneumonitis is rare (10%) but may be a life-threatening complication of these drugs. Pneumonitis symptoms include cough, shortness of breath, dyspnea, and fever, and often involve only asymptomatic radiographic changes. Pneumonitis shows ground-glass patchy lesions and/or disseminated nodular infiltrates, predominantly in the lower lobes. Identifying the exact cause of a pneumonitis in a patient treated with ICIs could be challenging during the current COVID-19 outbreak (**Fig. 75-5A**). Chest CT manifestations of COVID-19 include an imaging pattern of pure GGO, consolidation, nodules, fibrous stripes, and mixed patterns, with the distribution slightly predominant in the lower lobe and peripheral areas of the lung. Treatment of immune-mediated pneumonitis includes temporary or permanent withdrawal of drug and the addition of high-dose glucocorticoids (**Fig. 75-5B**).

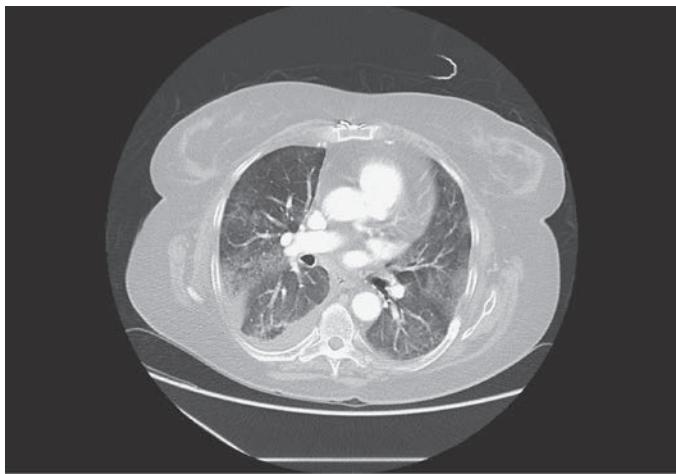
Radiation pneumonitis and/or fibrosis are relatively frequent side effects of thoracic radiation therapy. It may be acute or chronic. Radiation-induced lung toxicity is a function of the irradiated lung volume, dose per fraction, and radiation dose. The larger the irradiated lung field, the higher is the risk for radiation pneumonitis. The use of concurrent chemoradiation, particularly regimens including paclitaxel, increases pulmonary toxicity. Radiation pneumonitis usually develops 2–6 months after completion of radiotherapy. The clinical

syndrome, which varies in severity, consists of dyspnea, cough with scanty sputum, low-grade fever, and an initial hazy infiltrate on chest radiographs. The infiltrate and tissue damage usually are confined to the radiation field. The CT scan may show GGOs, consolidation, fibrosis, atelectatic cicatrization, pleural volume loss, or pleural thickening. The patients subsequently may develop a patchy alveolar infiltrate and air bronchograms, which may progress to acute respiratory failure that is sometimes fatal. A lung biopsy may be necessary to make the diagnosis. Asymptomatic infiltrates found incidentally after radiation therapy need not be treated. However, prednisone should be administered to patients with fever or other symptoms. The dosage should be tapered slowly after the resolution of radiation pneumonitis, because abrupt withdrawal of glucocorticoids may cause an exacerbation of pneumonia. Delayed radiation fibrosis may occur years after radiation therapy and is signaled by dyspnea on exertion. Often it is mild, but it can progress to chronic respiratory failure. Therapy is supportive.

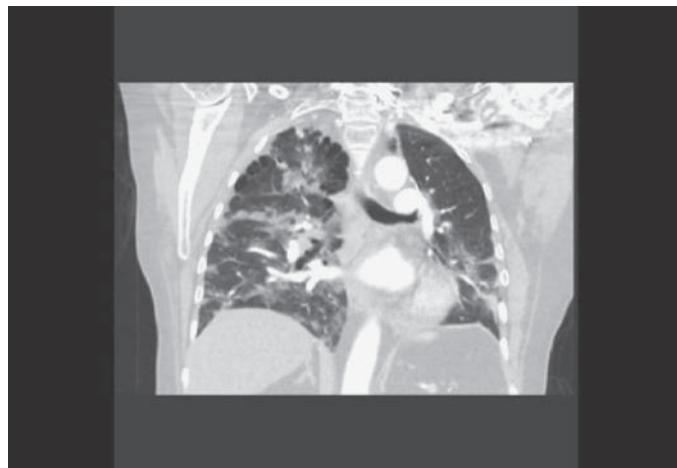
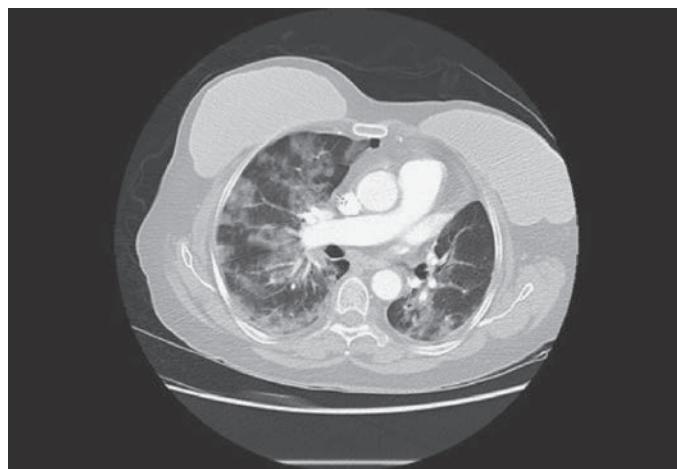
Classic radiation pneumonitis that leads to pulmonary fibrosis is due to radiation-induced production of local cytokines such as platelet-derived growth factor β , tumor necrosis factor, interleukins, and transforming growth factor β in the radiation field.

SBRT is a radiotherapy treatment method that has been applied to the treatment of stage I lung cancers in medically inoperable patients. SBRT accurately delivers a high dose of irradiation in one or few treatment fractions to an image-defined lung mass. Most of the acute changes after SBRT occur later than 3 months after treatment, and the shape of the SBRT-induced injury conforms more tightly to the tumor.

Pneumonia is a common problem in patients undergoing treatment for cancer (**Chap 74**). In patients with pulmonary infiltrates who are



A



B

FIGURE 75-5 **A.** Computed tomography scan of a 63-year-old female with metastatic adenocarcinoma on nivolumab with immune check point inhibitor pneumonia showing interlobular septal thickening and diffuse ground glass opacity to nivolumab. **B.** Computed tomography scan of a 68-year-old female with resected adenocarcinoma of lung and COVID 19 pneumonia showing peripheral and basilar predominant patchy groundglass and consolidative opacity consistent with multifocal COVID pneumonia.

afebrile, heart failure and multiple pulmonary emboli are in the differential diagnosis.

■ NEUTROGENIC ENTEROCOLITIS

Neutropenic enterocolitis (typhlitis) is the inflammation and necrosis of the cecum and surrounding tissues that may complicate the treatment of acute leukemia. Nevertheless, it may involve any segment of the gastrointestinal tract including small intestine, appendix, and colon. This complication has also been seen in patients with other forms of cancer treated with taxanes, 5-fluorouracil, irinotecan, vinorelbine, cisplatin, carboplatin, and high-dose chemotherapy (Fig. 75-6). It also has been reported in patients with AIDS, aplastic anemia, cyclic neutropenia, idiosyncratic drug reactions involving antibiotics, and immunosuppressive therapies. The patient develops right lower quadrant abdominal pain, often with rebound tenderness and a tense, distended abdomen, in a setting of fever and neutropenia. Watery diarrhea (often containing sloughed mucosa) and bacteremia are common, and bleeding may occur. Plain abdominal films are generally of little value in the diagnosis; CT scan may show marked bowel wall thickening, particularly in the cecum, with bowel wall edema, mesenteric stranding, and ascites, and may help to differentiate neutropenic colitis from other abdominal disorders such as appendicitis, diverticulitis, and *Clostridium difficile*-associated colitis in this high-risk population. Patients

with bowel wall thickness >10 mm on ultrasonogram have higher mortality rates. However, bowel wall thickening is significantly more prominent in patients with *C. difficile* colitis. Pneumatosis intestinalis is a more specific finding, seen only in those with neutropenic enterocolitis and ischemia. The combined involvement of the small and large bowel suggests a diagnosis of neutropenic enterocolitis. Rapid institution of broad-spectrum antibiotics, bowel rest, and nasogastric suction may reverse the process. Use of myeloid growth factors improved outcome significantly. Surgical intervention is reserved for severe cases of neutropenic enterocolitis with evidence of perforation, peritonitis, gangrenous bowel, or gastrointestinal hemorrhage despite correction of any coagulopathy.

C. difficile colitis is increasing in incidence. Newer strains of *C. difficile* produce ~20 times more of toxins A and B compared to previously studied strains. *C. difficile* risk is also increased with chemotherapy. Antibiotic coverage for *C. difficile* should be added if pseudomembranous colitis cannot be excluded.

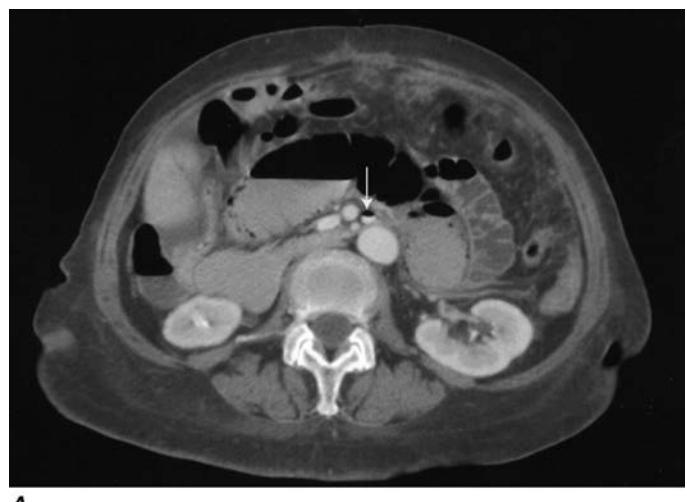
■ HEMORRHAGIC CYSTITIS

Hemorrhagic cystitis is characterized by diffuse bladder mucosal bleeding that develops secondary to chemotherapy (mostly cyclophosphamide or ifosfamide), radiation therapy, bone marrow transplantation (BMT), and/or opportunistic infections. Both cyclophosphamide and ifosfamide are metabolized to acrolein, which is a strong chemical irritant that is excreted in the urine. Prolonged contact or high concentrations may lead to bladder irritation and hemorrhage. Symptoms include gross hematuria, frequency, dysuria, burning, urgency, incontinence, and nocturia. The best management is prevention. Maintaining a high rate of urine flow minimizes exposure. In addition, 2-mercaptoethanesulfonate (mesna) detoxifies the metabolites and can be coadministered with the instigating drugs. Mesna usually is given three times on the day of ifosfamide administration in doses that are each 20% of the total ifosfamide dose. If hemorrhagic cystitis develops, the maintenance of a high urine flow may be sufficient supportive care. If conservative management is not effective, irrigation of the bladder with a 0.37–0.74% formalin solution for 10 min stops the bleeding in most cases. *N*-Acetylcysteine may also be an effective irrigant. Prostaglandin (carboprost) can inhibit the process. In extreme cases, ligation of the hypogastric arteries, urinary diversion, or cystectomy may be necessary.

In the BMT setting, early-onset hemorrhagic cystitis is related to drugs in the treatment regimen (e.g., cyclophosphamide), and late-onset hemorrhagic cystitis is usually due to the polyoma virus BKV or adenovirus type 11. BKV load in urine alone or in combination with acute graft-versus-host disease correlates with development of hemorrhagic cystitis. Viral causes are usually detected by polymerase chain reaction (PCR)-based diagnostic tests. Treatment of viral hemorrhagic cystitis is largely supportive, with reduction in doses of immunosuppressive agents, if possible. No antiviral therapy is approved, although cidofovir was reported to be effective in a small series. Hyperbaric oxygen therapy has been used successfully in patients with BKV-associated and cyclophosphamide-induced hemorrhagic cystitis during hematopoietic stem cell transplantation, as well as in hemorrhagic radiation cystitis that occurs in up to 5% of patients after pelvic radiation.

■ HYPERSENSITIVITY REACTIONS TO ANTINEOPLASTIC DRUGS

Many antineoplastic drugs may cause hypersensitivity reaction. These reactions are unpredictable and potentially life threatening. Most reactions occur during or within hours of parenteral drug administration. Taxanes, platinum compounds, asparaginase, etoposide, procarbazine, and biologic agents, including rituximab, bevacizumab, trastuzumab, gemtuzumab, cetuximab, and alemtuzumab, are more commonly associated with acute hypersensitivity reactions than are other agents. Hypersensitivity reactions to some drugs, such as taxanes, occur during the first or second dose administered. Hypersensitivity to platinum compounds occurs after prolonged exposure. Skin testing may identify patients with high risk for hypersensitivity after carboplatin exposure. Premedication with histamine H_1 and H_2 receptor antagonists and glucocorticoids reduces the incidence of hypersensitivity reaction to



A



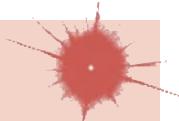
B

FIGURE 75-6 Abdominal computed tomography (CT) scans of a 72-year-old woman with neutropenic enterocolitis secondary to chemotherapy. **A.** Air in inferior mesenteric vein (arrow) and bowel wall with pneumatosis intestinalis. **B.** CT scan of upper abdomen demonstrating air in portal vein (arrows).

76

Cancer of the Skin

Brendan D. Curti, John T. Vetto,
Sancy A. Leachman



MELANOMA

Pigmented lesions are among the most common findings on skin examination. The challenge for the physician is to distinguish benign lesions from cutaneous melanomas and nonmelanoma skin cancers (NMSCs). Both melanoma and NMSC are increasing in frequency, and melanoma accounts for over half of the deaths resulting from skin cancer. Melanoma is an aggressive malignancy of melanocytes, pigment-producing cells that originate from the neural crest and migrate to the skin, meninges, mucous membranes, upper esophagus, and eyes. Melanocytes in each of these locations have the potential for malignant transformation, but the vast majority of melanomas arise in the skin, often permitting detection at a time when complete surgical excision leads to cure. Cutaneous melanoma can occur in people of all ages and all colors. Examples of malignant melanoma of the skin, mucosa, eye, and nail are shown in **Fig. 76-1**.

taxanes, particularly paclitaxel. Despite premedication, hypersensitivity reactions may still occur. In these cases, rapid desensitization in the intensive care unit setting or re-treatment may be attempted with care, but the use of alternative agents may be required. Skin testing is used to assess the involvement of IgE in the reaction. Tryptase levels measured at the time of the reaction help to explain the mechanism of the reaction and its severity. Increased tryptase levels indicate underlying mast cell activation. Candidate patients for desensitization include those who have mild to severe hypersensitivity type I, with mast cell-mediated and IgE-dependent reactions occurring during a chemotherapy infusion or shortly thereafter.

FURTHER READING

- AZIZI AH et al: Superior vena cava syndrome. *JACC Cardiovasc Interv* 13:2896, 2020.
- CASTELLS M et al: Hypersensitivity to antineoplastic agents: Mechanisms and treatment with rapid desensitization. *Cancer Immunol Immunother* 61:1575, 2012.
- CASTINETTI F et al: Endocrine side-effects of new anticancer therapies: Overall monitoring and conclusions. *Ann Endocrinol (Paris)* 79:591, 2018.
- DURANI U, HOGAN WJ: Emergencies in haematology: Tumour lysis syndrome. *Br J Haematol* 188:494, 2020.
- FAJGENBAUM DC, JUNE CH: Cytokine storm. *N Engl J Med* 383:2255, 2020.
- GONZALEZ CASTRO LN, MILLIGAN TA: Seizures in patients with cancer. *Cancer* 126:1379, 2020.
- LAWTON AJ et al: Assessment and management of patients with metastatic spinal cord compression: A multidisciplinary review. *J Clin Oncol* 37:61, 2019.
- LIN AL, AVILA EK: Neurologic emergencies in the patients with cancer. *J Intensive Care Med* 32:99, 2017.
- PELLERINO A et al: Neoplastic meningitis in solid tumors: From diagnosis to personalized treatments. *Ther Adv Neurol Disord* 11:1756286418759618, 2018.
- RIAZ A et al: Percutaneous management of malignant biliary disease. *J Surg Oncol* 120:45, 2019.
- SCHUSLER R, MEYERSON SL: Pericardial disease associated with malignancy. *Curr Cardiol Rep* 20:92, 2018.
- THOMAS MR, SCULLY M: How I treat microangiopathic hemolytic anemia in patients with cancer. *Blood* 137:1310, 2021.

RISK FACTORS AND EPIDEMIOLOGY

The epidemiologic patterns seen in melanoma reflect the genetic and biologic features of melanocytes and their response to environmental ultraviolet radiation (UVR). Clinical features that confer an increased risk for melanoma include: (1) vulnerability to sun damage (light/red coloration of skin, hair, or eyes; photodamaged skin; history of exposure to natural or artificial UVR; prior history of skin cancers of any type); (2) abnormal growth of melanocytes (increased absolute number of nevi, increased size of nevi, or atypical features of moles such as multiple colors, speckles, or shapes); and (3) immunosuppression (innate, functional, or drug-induced). **Table 76-1** summarizes melanoma risk factors and the relative risk associated with these factors.

The incidence and mortality rates are strongly influenced by ethnic and geographic/environmental factors, which superimpose substantial variability on melanoma rates. Specifically, the incidence of melanoma is 1/100,000 per year in populations with high skin eumelanin content and up to 27/100,000 per year in populations with low skin eumelanin. Men are affected slightly more than women (1.3:1), and the median age at diagnosis is the late fifties. Melanoma is one of the few cancer types with increasing incidence in the United States and is now the fifth leading cancer in men (60,190 new cases estimated in 2020; probability 1:28) and the sixth leading cancer in women (40,160 new cases estimated in 2020; probability 1:41). Although these rankings are based on the total number of new invasive melanoma cases (100,350 in 2020), an additional 95,710 cases of melanoma in situ are expected to occur in 2020. Given the stable or decreasing mortality (see below), it seems likely that new cases include some that represent overdiagnosis of cancers that would not progress to fatal disease.

Mortality rates begin to rise at age 55, with the greatest mortality in men age >65 years. In contrast to the increasing incidence, the mortality rates for melanoma are decreasing, though this trend appears less dramatic outside of the United States. The reasons for the decrease are not entirely clear but have been attributed in part to the recent success of melanoma therapeutics on survival. After U.S. Food and Drug Administration (FDA) approval of ipilimumab and vemurafenib in 2011, the 1-year relative survival rate increased from 42% (2008–2010) to 55% (2013–2015). The mortality rate from 2013 to 2017 dropped annually by 7% in those aged 20–64 years old and dropped 5–6% per year for individuals aged ≥65 years.

GLOBAL CONSIDERATIONS

The incidence of both nonmelanoma and melanoma skin cancers around the world has been increasing. Every year, between 2 and 3 million people develop NMSC, and in 2018, there were 300,000 cases of melanoma. A disproportionate number of cases and deaths occur in North America, Europe, Australia, and New Zealand. The highly variable incidence rates of melanoma in different populations are due to the interplay between risk factors, including host genetics and environmental factors, that distribute risk unevenly across these populations and account for the absolute risk in different ethnic groups and geographic areas.

Dark-skinned populations (such as those of India and Puerto Rico), blacks, and East Asians also develop melanoma but at rates 10–20 times lower than those in whites. Cutaneous melanomas in dark-skinned populations are more often diagnosed at a higher stage, and patients tend to have worse outcomes. Surveillance, Epidemiology, and End Results (SEER) data (2000–2004) reveal that whites have the highest incidence of melanoma at 27.2/100,000 and that the incidence drops substantially in Hispanics (4.5/100,000), Native Americans (4.1/100,000), Asians/Pacific Islanders (1.7/100,000), and blacks (1.1/100,000). In nonwhite populations, the frequency of acral (subungual, plantar, palmar) and mucosal melanomas is much higher; the incidence of melanoma in black and Hispanic populations is not associated with ultraviolet (UV) exposure. In China, about 20,000 new melanomas are reported each year, and in contrast to the United States, mortality is increasing. This may be due to the fact that in Asians and dark-skinned populations, more melanomas arise from acral and mucosal areas, which have a different biology and carry a poorer



FIGURE 76-1 Types of melanoma. **A.** Hypomelanotic melanoma. **B.** Superficial spreading melanoma. **C.** Melanoma arising in a nevus. **D.** Melanoma arising in a nevus. **E.** Nodular melanoma. **F.** Cutaneous melanoma metastases at a surgical margin (also known as melanoma satellites when <2 cm from the primary tumor and in-transit melanoma when >2 cm). **G.** Mucosal melanoma arising in the vulva. **H.** Choroidal melanoma with tumor borders marked by arrowheads, color fundus photograph. **I.** Acral melanoma with Hutchinson's sign on the proximal nail fold.

prognosis than cutaneous melanomas. Little is yet known about the effects of mixed ethnicity on melanoma risk.

■ GENETIC SUSCEPTIBILITY TO MELANOMA

Approximately 20–40% of hereditary melanomas (0.2–2% of all melanomas) are due to germline mutations in the cell cycle regulatory gene cyclin-dependent kinase inhibitor 2A (*CDKN2A*). In fact, 70% of all cutaneous melanomas have mutations or deletions affecting the *CDKN2A* locus on chromosome 9p21. This locus encodes two distinct tumor-suppressor proteins from alternate reading frames: p16 and ARF (p14^{ARF}). The p16 protein inhibits CDK4/6-mediated phosphorylation and inactivation of the retinoblastoma (RB) protein, whereas ARF inhibits MDM2 ubiquitin-mediated degradation of p53. The loss of *CDKN2A* results in inactivation of two critical tumor-suppressor pathways, RB and p53, which control entry of cells into the cell cycle. Several studies have shown an increased risk of pancreatic cancer among melanoma-prone families with *CDKN2A* mutations. A second high-risk locus for melanoma susceptibility, *CDK4*, is located on chromosome 12q13 and encodes the cyclin-dependent kinase inhibited by p16. *CDK4* mutations, which also inactivate the RB pathway, are much rarer than *CDKN2A* mutations. Germline mutations in the melanoma

lineage-specific oncogene microphthalmia-associated transcription factor (*MITF*), BRCA1-associated protein 1 (*BAP-1*), protection of telomeres 1 (*POT-1*), and telomerase reverse transcriptase (*TERT*) also predispose to familial melanoma with a not yet quantified high penetrance, based on families that have been tested.

The melanocortin-1 receptor (*MC1R*) gene is a moderate-risk inherited melanoma susceptibility factor. UVR stimulates the production of melanocortin (α -melanocyte-stimulating hormone [α -MSH]), the ligand for *MC1R*, which is a G-protein-coupled receptor that signals via cyclic AMP and regulates the amount and type of pigment produced by melanocytes. *MC1R* is highly polymorphic, and many among its ~80 variants result in partial or full loss of signaling and lead to the production of non-photoprotective red/yellow pheomelanins, rather than photoprotective brown/black eumelanins. The red hair color (RHC) phenotype produced by *MC1R* mutations includes lightly colored skin, red hair, freckles, increased sun sensitivity, and increased risk of melanoma. In addition to its weak UV-shielding capacity relative to eumelanin, increased pheomelanin production in patients with inactivating polymorphisms of *MC1R* also provides a UV-independent carcinogenic contribution to melanogenesis via oxidative damage and reduced DNA damage repair.

TABLE 76-1 Melanoma Risk Factors and Relative Risk

RISK LEVEL	RISK FACTOR	RELATIVE RISK
Elevated	1 atypical nevus versus 0	1.5
	Total common nevi, 16+ versus <15	1.5
	Blue eye color versus dark	1.5
	Hazel eye color versus dark	1.5
	Green eye color versus dark	1.6
	Light brown hair versus dark	1.6
	Indoor tanning in any gender versus never	1.7
	Fitzpatrick II versus IV	1.8
	Fitzpatrick III versus IV	1.8
	History of sunburn versus no sunburn	2.0
	Blond hair versus dark	2.0
	2 atypical nevi versus 0	2.1
	Fitzpatrick I versus IV	2.1
	High density of freckles versus none	2.1
	Total common nevi 41–60 versus <15	2.2
Moderately elevated	Family history of melanoma in 1 or more first-degree relatives	1.7–3.0
	3 atypical nevi versus 0	3.0
	Total common nevi 61–80 versus <15	3.3
	Red hair versus dark	3.6
	Chronic lymphocytic leukemia	3.9
	History of actinic keratoses and/or keratinocyte carcinoma versus not	4.3
	Indoor tanning in women aged 30–39 versus never	4.3
High	4 atypical nevi versus 0	4.4
	Transplant recipient versus not	2.2–4.6
	Indoor tanning in women aged <30 versus never	6.0
	5 atypical nevi versus 0	6.4
	Total common nevi 81–120 versus <15	6.9
	Personal history of melanoma	8.2–13.4
	CDKN2A mutation carrier	14–28

Other more common, low-penetrance polymorphisms in genes related to pigmentation, nevus count, immune responses, DNA repair, metabolism, and the vitamin D receptor have small effects on melanoma susceptibility. In sum, ~50–60% of the genetic risk for hereditary melanoma can be attributed to known melanoma predisposition genes, with ~40% of the known genetic risk attributable to *CDKN2A*. The other components of inherited risk are most likely due to the presence of additional modifier genes and/or shared environmental exposures of the host.

■ PREVENTION AND EARLY DETECTION

Primary prevention of melanoma and NMSC is based on protection from the sun. Public health initiatives, such as the SunSmart program that started in Australia and is now operative in Europe and the United States, have demonstrated that behavioral change can decrease the incidence of NMSC and melanoma. Preventive measures should start early in life because damage from UV light begins early despite the fact that cancers develop years later. Early episodes of sun burns may be a greater risk than chronic tanning. Some individuals tan compulsively. There is greater understanding of tanning addiction and the cutaneous-neural connections that may give rise to this behavior. Compulsive tanners exhibit differences in dopamine binding and reactivity in reward pathways in the brain, such as the basal striatum, resulting in cutaneous secretion of β -endorphins after UV exposure. Identifying individuals with tanning addiction may be another prevention method. Regular use of broad-spectrum sunscreens that block UVA and UVB

with a sun protection factor (SPF) of at least 30 and protective clothing should be encouraged. Physical blockers such as zinc oxide and titanium dioxide have less likelihood of being absorbed or of generating an allergic reaction than chemical sunscreens. Avoidance of sunburns, tanning beds, and midday sun exposure is recommended.

Secondary prevention comprises education and screening with the goal of early detection and can be individualized based on risk factors. A full-body skin exam is warranted in populations at higher risk for melanoma such as patients with clinically atypical moles (dysplastic nevi) and those with a personal history of melanoma. Surveillance in high-risk patients should be performed by a dermatologist and include total-body photography and dermoscopy where appropriate. Individuals with three or more primary melanomas and families with at least one invasive melanoma and two or more cases of melanoma and/or pancreatic cancer among first- or second-degree relatives on the same side of the family may benefit from genetic testing. Severely atypical nevi and melanoma in situ should be removed. Early detection of small lesions allows the use of simpler treatment modalities with higher cure rates and lower morbidity. Monthly self-screening augments provider-based screening. Patients should be taught to recognize the clinical features of melanoma and advised to report any change in a pigmented lesion. There is evidence supporting the ability of media campaigns to reduce cancer mortality in lung cancer, and results from Australia's skin cancer campaigns demonstrate improvement in attitude and behavior and a reduction in melanoma incidence. A benefit/cost analysis in Australia showed a return of \$3.85 for every \$1 invested. Although the U.S. Preventive Services Task Force states that there is insufficient evidence to recommend skin screening for the general population, additional research is anticipated to find best practices for skin cancer detection and prevention.

■ DIAGNOSIS

The goal is to identify a melanoma before it becomes invasive and life-threatening metastases have occurred. Early detection may be facilitated by applying the ABCDEs: *asymmetry* (benign lesions are usually symmetric); *border irregularity* (most nevi have clear-cut borders); *color variegation* (benign lesions usually have uniform light or dark pigment); *diameter >6 mm* (the size of a pencil eraser); and *evolving* (any change in size, shape, color, or elevation or new symptoms such as bleeding, itching, and crusting). In addition, any nevus that appears atypical and different from the rest of the nevi on that individual (an “ugly duckling”) should be considered suspicious.

The entire skin surface, including the scalp and mucous membranes, as well as the nails should be examined in each patient. Bright room illumination is important, and a hand lens or dermatoscope is helpful for evaluating variation in pigment pattern. Any suspicious lesions should be biopsied, evaluated by a specialist, or recorded by chart and/or photography for follow-up. Dermoscopy employs low-level magnification of the epidermis with polarized light or water interface and permits a more precise visualization of patterns of pigmentation than is possible with the naked eye (Fig. 76-2).

Biopsy Any pigmented cutaneous lesion that has changed in size or shape or has other features suggestive of malignant melanoma is a candidate for biopsy. An excisional biopsy with 1- to 3-mm margins (narrow-margin excision) is suggested. This facilitates histologic assessment of the lesion, permits accurate measurement of thickness if the lesion is melanoma, and constitutes definitive treatment if the lesion is benign. For lesions that are large or on anatomic sites where excisional biopsy may not be feasible (such as the face, hands, and feet), an incisional biopsy (partial biopsy) through the most nodular or darkest area of the lesion is acceptable. Incisional biopsy does not appear to facilitate the spread of melanoma. For suspicious lesions, every attempt should be made to preserve the ability to assess the deep and peripheral margins and to perform immunohistochemistry. Shave, saucerization, or punch biopsies are an acceptable alternative, particularly if the suspicion of malignancy is low. They should be deep enough to include the deepest component of the entire lesion, and any pigment at the base of the lesion should be removed and included with the biopsy specimen.

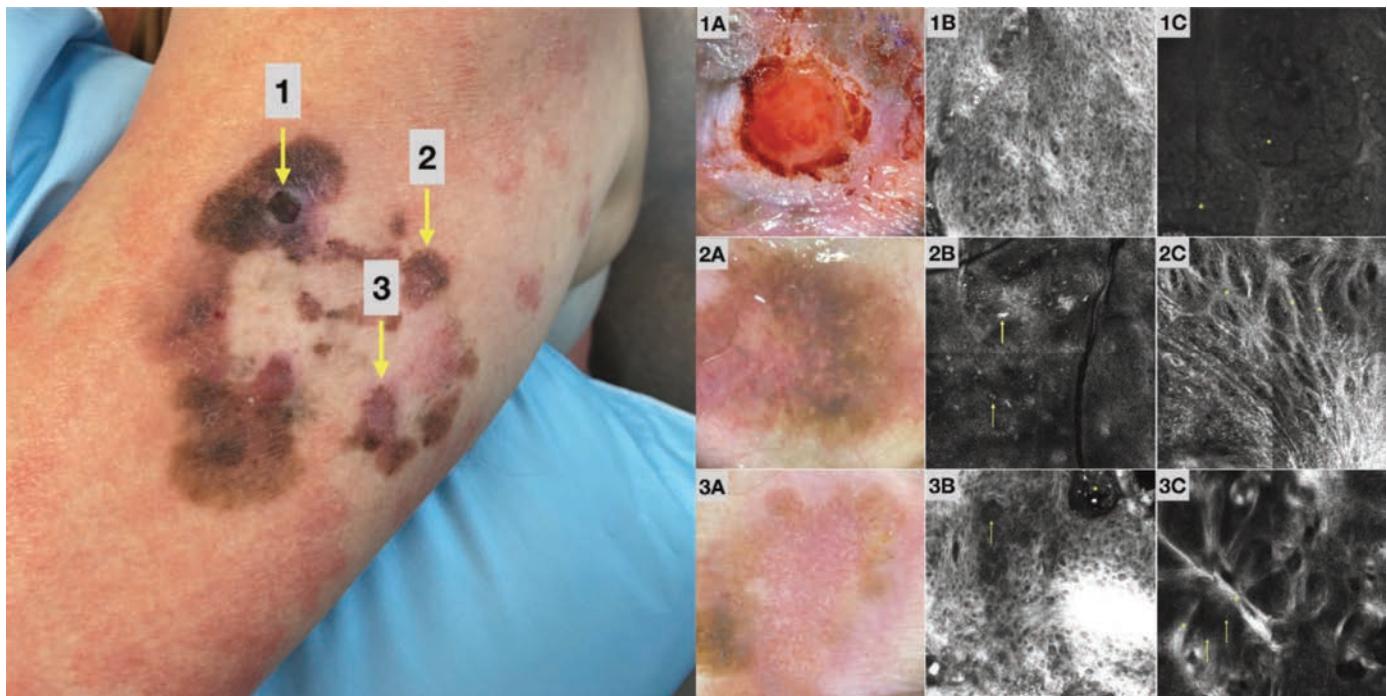


FIGURE 76-2 Clinical and confocal diagnostic findings of melanoma. **Left panel:** A clinical image of a large melanoma of a 60-year-old woman used to illustrate classic features of nodular melanoma (1), superficial spreading melanoma (2), and amelanotic melanoma (3). Panels **1A**, **2A**, and **3A** correspond to the dermoscopy images taken at sites 1, 2, and 3, respectively (Sklip Dermatoscope, Sklip LLC, Las Vegas, NV). Panels **1B**, **1C**, **2B**, **2C**, **3B**, and **3C** are reflectance confocal microscopy images of the epidermis and upper dermis taken at sites 1, 2, and 3, respectively (Vivascope 1500 Gen 4, Caliber I.D., Rochester, NY). **1A.** Site 1 dermoscopy shows a pink nodule with polymorphous vessels and ulceration with active bleeding consistent with malignancy. **1B.** Site 1 reflectance confocal microscopy of the epidermis shows an atypical enlarged honeycombed pattern frequently seen in melanoma. **1C.** Site 1 reflectance confocal microscopy of the upper dermis shows cerebriform nests (*) seen in nodular melanoma. **2A.** Site 2 dermoscopy shows a pigmented area with an atypical, thickened network, blue-gray structures, and polymorphous vessels. **2B.** Site 2 reflectance confocal microscopy of the epidermis shows an irregular honeycombed pattern and pagetoid cells with nuclei (↑) typically seen in a superficial spreading melanoma. **2C.** Site 2 reflectance confocal microscopy of the dermoepidermal junction shows thickened junctional nests with bright reflective linear dendritic cells (*). **3A.** Site 3 dermoscopy shows an amelanotic area within the melanoma with milky red areas, polymorphous vessels, atypical network, and blue-gray structures classic for an amelanotic melanoma. **3B.** Site 3 reflectance confocal microscopy of the epidermis shows an irregular enlarged honeycombed pattern, dermal nests protruding upward into the epidermis (↑), and artefacts (*). **3C.** Site 3 reflectance confocal microscopy image of the dermoepidermal junction shows thickened collagen bundles (*) with atypical polymorphous vessels (↑).

Punch biopsies are more likely to clear the deep margin but more likely to be positive at the radial margins; the opposite is true for shave biopsies. The choice of biopsy type should be guided by which is most likely to remove the entire lesion for histologic evaluation.

The biopsy should be read by a pathologist experienced in pigmented lesions, and the report should include Breslow thickness, mitotic rate, presence or absence of ulceration, lymphatic invasion, regression, microsatellitosis, and the status of the peripheral and deep margins. Breslow thickness is the greatest thickness of a primary cutaneous melanoma measured on the slide from the top of the epidermal granular layer, or from the ulcer base, to the bottom of the tumor. To

distinguish melanomas from benign nevi in challenging cases, fluorescence *in situ* hybridization with multiple probes or comparative genomic hybridization can be helpful. Gene expression profile (GEP) assays have been developed to determine prognosis and are commercially available.

CLASSIFICATION AND PATHOGENESIS

Clinical Five major types of cutaneous melanoma are described in Table 76-2. In *superficial spreading melanoma*, *lentigo maligna melanoma*, and *acral lentiginous melanoma*, the lesion has a period of

TABLE 76-2 Major Histologic Subtypes of Malignant Melanoma

TYPE	SITE	APPEARANCE	ASSOCIATED MUTATIONS
Lentigo maligna	Sun-exposed surfaces, particularly malar region and temple	In flat portions, brown and tan predominate, but whitish gray sometimes present; in nodules, reddish brown, bluish gray, bluish black.	BRAF 28% NRAS 15%
Superficial spreading	Any (more common on upper back and, in women, lower legs)	Brown mixed with bluish red, bluish black, reddish brown, and often whitish pink. The lesion border is often visibly and/or palpably raised.	BRAF 57% NRAS 18%
Nodular	Any	Reddish blue, purple, or bluish black; can be uniform or mixed with brown and black.	BRAF 47% NRAS 33%
Acral lentiginous	Palm, sole, nail bed, mucous membrane	In flat portions, dark brown; in raised lesions (plaques), brown-black or blue-black.	NRAS 25% c-KIT 10% BRAF 10%
Desmoplastic	Any (more common on head and neck)	Highly variable; pigmentation is frequently absent. Can mimic nodular basal cell carcinoma.	MAPK and PI3K 73% High tumor mutational burden, BRAF and NRAS uncommon

Driver Mutations

BRAF: 10%
NRAS: 10%
C-KIT: 5–10%
NF1: 48% of *BRAF* and *NRAS* WT melanoma in older patients
BRAF: 50%
NRAS: 20%
C-KIT: 0%

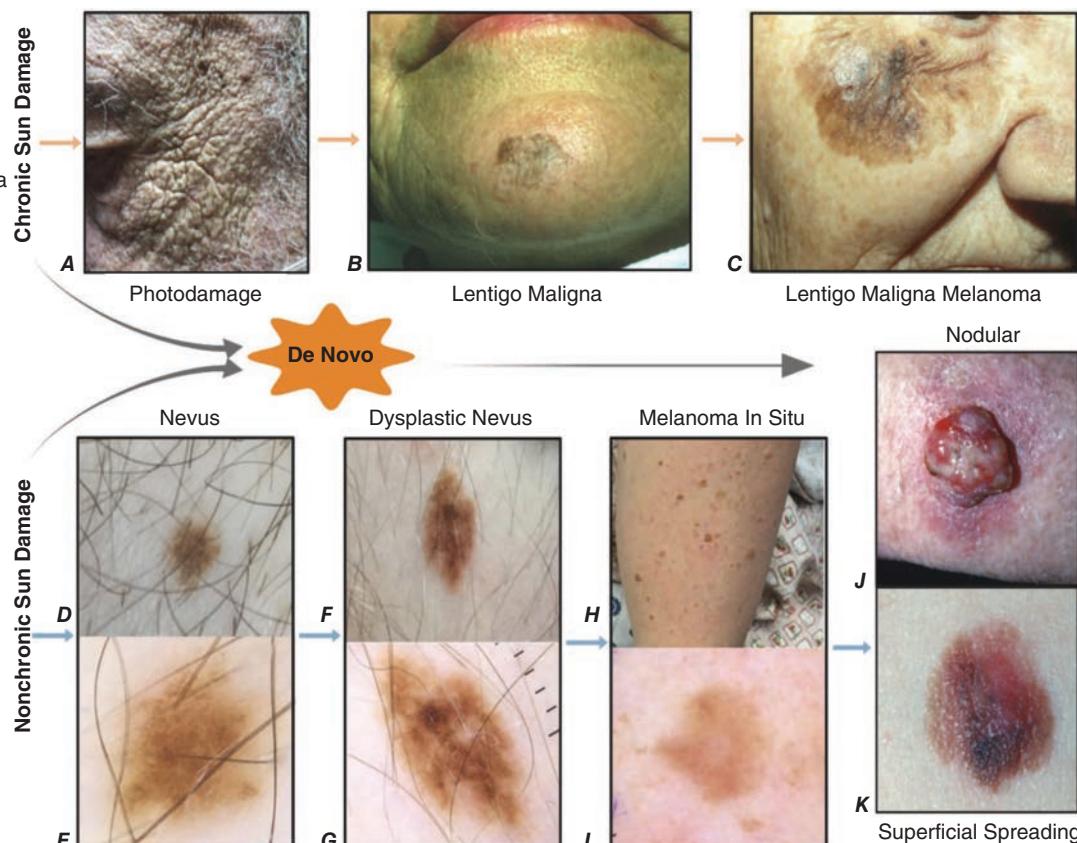


FIGURE 76-3 Cutaneous melanoma development and associated driver mutations. Chronic sun damage (**A**) predisposes to a lentigo maligna (in situ) (**B**), which can evolve into lentigo maligna melanoma (invasive) (**C**). Similarly, nonchronic sun damage can initiate melanoma de novo or in nevomelanocytes, where clinical and histologic changes of atypia may be seen prior to complete transformation. Nevi (**D**, **E**) can evolve into atypical lesions (**F**, **G**), in situ melanoma (**H**, **I**), and eventually invasive nodular (**J**) or superficial spreading melanomas (**K**).

superficial (radial) growth during which it increases in surface area but does not penetrate deeply and is most capable of being cured by surgical excision. Melanomas with a radial growth phase are characterized by irregular and sometimes notched borders, variation in pigment pattern, and variation in color. *Nodular melanoma* does not have a radial growth phase but usually presents with penetration deep into the skin (vertical growth phase). *Desmoplastic melanoma* is associated with a fibrotic response, neural invasion, and a greater tendency for local recurrence. Occasionally, melanomas appear clinically to be amelanotic (not pigmented), in which case the diagnosis is established microscopically after biopsy.

Although these subtypes are clinically distinct, they are primarily of historical interest because this classification has minimal prognostic value and is not part of American Joint Committee on Cancer (AJCC) staging. Characterizing the genomic and mutational profiles of melanoma has become increasingly common, informs prognosis, reflects the mechanisms of tumorigenesis, and may influence surveillance strategies and treatment.

Genomic The advent of next-generation sequencing has led to whole exome sequencing of hundreds of cutaneous melanomas derived from nonglabrous skin. This has revealed very complex genomic changes resulting from both germline (see “Genetic Susceptibility to Melanoma” above) and somatic mutations. Cutaneous melanomas have one of the highest somatic mutation rates (>10 mutations/Mb) among all cancers; the majority (76% of primary tumors and 84% of metastatic melanomas) exhibit mutations indicative of UVR exposure. The mutation rate varies based on body site; melanomas arising in chronic sun-damaged skin harbor substantially more mutations than melanomas from non-sun-damaged skin.

Melanomas can harbor thousands of mutations, but only a few are “driver” mutations that promote cell proliferation or inhibit normal

pathways of apoptosis or DNA repair and confer a growth advantage to the neoplastic cell. Some of the driver mutations for cutaneous melanoma are depicted in Fig. 76-3 along with the clinical evolution of melanoma lesions. Driver mutations are often found in combination with mutations to germline susceptibility genes such as *p16*, which affect cell cycle arrest, and *ARF*, which result in defective apoptotic responses to genotoxic damage. The altered melanocytes accumulate DNA damage and develop the malignant phenotype characterized by invasion, metastasis, and angiogenesis.

A genomic classification of cutaneous melanoma has been proposed based on the pattern of the most prevalent mutated genes, *BRAF*, *RAS*, and *NF1*, along with a triple wild-type (WT) in which no mutations in these three genes are found. The pattern of DNA mutations can vary with the site of origin and is independent of the histologic subtype of the tumor. An important aspect of this classification is that the mutational profile can guide therapy. The proliferative pathways affected by the mutations include the mitogen-activated protein (MAP) kinase and phosphatidylinositol 3' kinase/AKT pathways. *RAS* and *BRAF*, members of the MAP kinase pathway, which mediate the transcription of genes involved in cell proliferation and survival, undergo somatic mutation in melanoma and thereby generate potential therapeutic targets. *NRAS* is mutated in ~20% of melanomas, and somatic activating *BRAF* mutations are found in most benign nevi and 40–50% of cutaneous melanomas. Neither mutation by itself appears to be sufficient to cause melanoma; thus, they often are accompanied by other mutations, such as *TERT*. The *BRAF* mutation is most commonly a T→A point mutation that results in a valine-to-glutamate amino acid substitution (V600E). V600E *BRAF* mutations are more common in younger patients and are present in most melanomas that arise on sites with intermittent sun exposure and are less common in melanomas from chronically sun-damaged skin (i.e., those of older patients).

Melanomas may harbor mutations in *AKT* (primarily in *AKT3*) and *PTEN* (phosphatase and tensin homolog). *AKT* can be amplified, and

PTEN may be deleted or undergo epigenetic silencing that leads to constitutive activation of the PI3K/AKT pathway and enhanced cell survival by antagonizing the intrinsic pathway of apoptosis. A loss-of-function mutation in *NF1*, which can affect both the MAP kinase and PI3K/AKT pathways, has been described in 10–15% of melanomas. In melanoma, these two signaling pathways (MAP kinase and PI3K/AKT) enhance tumorigenesis, chemoresistance, migration, and cell cycle dysregulation.

■ PROGNOSTIC FACTORS

The most important prognostic factors for a newly diagnosed patient are incorporated in the staging classification. The best predictor of recurrence is Breslow thickness, followed by ulceration, which together make up the T stage of the AJCC system for melanoma. The anatomic site of the primary tumor is also prognostic; favorable sites are the forearm and leg, and unfavorable sites include the scalp, hands, feet, and mucous membranes. Women with stage I or II disease have better survival than men, perhaps in part because of earlier diagnosis; women frequently have melanomas on the lower leg, where self-recognition is more likely and the prognosis is better.

Older individuals, especially men >60, have a tendency toward delayed diagnosis (and thus thicker tumors), have more head and neck and acral melanomas (which tend to have earlier vertical growth and distant metastases), and are more likely to develop melanomas in chronically UVR-damaged skin (which are more often *BRAF* wild type, with fewer options for therapy). Other important adverse factors include high mitotic rate, microscopic evidence of regression, and lymphatic/vascular invasion. Clinical features such as microsatellite lesions and/or in-transit metastases, evidence of nodal involvement, elevated serum lactate dehydrogenase (LDH), and presence and site of distant metastases all portend a higher stage and worse prognosis.

GEPs and machine-learning algorithms that associate genomic changes with clinical outcomes have been used to estimate the prognosis of melanoma. A commercially available 31-gene GEP is available that predicts for all-site (particularly distant) relapse and incorporates the increased and decreased expression, as well as the dysregulation, of genes involved in many of the cellular processes leading to melanoma progression described earlier. Although this 31-gene GEP can estimate the probability of distant relapse, it has not supplanted the prognostic estimates derived from surgical staging. It is anticipated that GEPs will be incorporated into future cutaneous melanoma management guidelines, as they have been for uveal melanoma, breast cancer, thyroid cancer, and other malignancies.

■ STAGING

Once the diagnosis of melanoma has been made, the tumor is staged to determine the prognosis and aid in treatment selection. The current melanoma staging criteria and estimated 10-year survival by stage are depicted in **Table 76-3**. The clinical stage is determined after the microscopic evaluation of the melanoma skin lesion and clinical and radiologic assessment. The pathologic stage also includes microscopic examination of clinically negative regional lymph nodes obtained at sentinel lymph node biopsy (SLNB), any enlarged nodes found on exam or imaging, and any suspected metastases amenable to open or image-guided biopsy.

All patients should have a complete history, with attention to symptoms that suggest metastatic disease, such as new palpable masses, malaise, weight loss, headaches, changes in vision or bowel habits, hemoptysis, and pain. The provider should look for persistent disease at the biopsy site, dermal or subcutaneous nodules that could represent satellite or in-transit metastases, and lymphadenopathy. A complete blood count, complete metabolic panel, and LDH should be performed. Although these tests rarely help uncover occult metastatic disease, a microcytic anemia would raise the possibility of bowel metastases, elevated liver function tests can suggest liver metastases, and LDH is part of the AJCC system for stage IV disease. Abnormal test results should prompt a more extensive evaluation, including computed tomography (CT) scan or a positron emission tomography (PET) scan (or CT/PET combined).

TABLE 76-3 Staging and Survival

STAGE	TNM	10-YEAR MELANOMA-SPECIFIC SURVIVAL ESTIMATE
0	TisN0M0	>99%
IA	T1aN0M0, T1bN0M0	98–96%
IB	T2aN0M0	92%
IIA	T2b-T3aN0M0	88%
IIB	T3b-T4aN0M0	81–83%
IIC	T4bN0M0	75%
IIIA	T1a-T2aN1a-2aM0	88%
IIIB	T2b-T3aN1a-N2bM0	77%
IIIC	T3b-4bN1a-N3cM0	60%
IIID	T4bN3a-N3cM0	24%
IV M1a	Any T, any N, skin, soft tissue, or distant nodal sites	50% at 5 years
IV M1b	Any T, any N, lung + any M1a sites	35–50% at 5 years
IV M1c	Any T, any N, skin, non-CNS visceral disease, any M1a or M1b sites	~25% at 5 years
IV M1d	Any T, any N, CNS metastasis + any M1a,b,c sites	<5% at 5 years

Abbreviations: CNS, central nervous system; TNM, tumor-node-metastasis.

Despite all the above considerations, >80% of patients at presentation will have disease confined to the skin and a negative history and physical examination, in which case imaging is not indicated. Although controversial, an exception is sometimes made for very-high-risk primaries (e.g., >4 mm with ulceration) in which the chance for occult distant metastases is higher than that for a positive SLNB.

TREATMENT

Melanoma

MANAGEMENT OF CLINICALLY LOCALIZED MELANOMA (STAGE I, II)

For a newly diagnosed cutaneous melanoma, wide surgical excision of the lesion with a margin of normal skin is necessary to remove all malignant cells and minimize the probability of local recurrence. The National Comprehensive Cancer Network (NCCN), based on data from six randomized trials, recommends the following radial margins for a primary melanoma: *in situ*, 0.5–1.0 cm; invasive up to 1 mm thick, 1 cm; >1.01–2 mm, 1–2 cm; and >2 mm, 2 cm. Smaller margins may be used for special locations such as the face, hands, feet, and genitalia due to the higher likelihood of morbidity in these regions. In all instances, however, inclusion of subcutaneous fat in the surgical specimen facilitates adequate thickness measurement and assessment of surgical margins by the pathologist. When feasible, excision should go down to fascia, with fascial resection for thick (T4) lesions. Topical imiquimod can be used to treat lentigo maligna in cosmetically sensitive locations with narrow resection margins by promoting local immune response resulting in decreased local recurrence.

SLNB is a valuable staging tool providing prognostic information to identify patients at high risk for relapse who may be candidates for adjuvant therapy. The first (sentinel) draining node(s) from the primary site is (are) located by injecting a blue dye and a gamma-emitting radioisotope around the primary site. The sentinel node(s) then is (are) identified using a handheld gamma detector brought sterilely into the operative field. The surgeon makes an incision of the area of uptake and looks for the blue-stained, “hot” node(s), which is (are) removed and subjected to histopathologic analysis with serial sectioning using hematoxylin and eosin and immunohistochemical stains (e.g., S100, HMB45, MART-1, and MelanA) to identify melanocytes.

NCCN guidelines recommend SLNB to patients with a 10% or greater chance of having tumor in the node. This includes patients

with tumors >1 mm thick (T2) or T1 tumors that have ulceration (T1b). Patients with a 5–10% risk of node positivity, such as those with tumors measuring between 0.75 and 1.0 mm, transected tumors, regressed tumors, or lymphovascular invasion, should also be considered for SLNB. The NCCN does not recommend SLNB for patients with a risk of a positive SLNB ≤5% such as those with melanomas ≤0.75 mm thick and no high-risk features. In these patients, wide excision alone is the usual definitive therapy.

Patients whose SLNB is negative can either be followed or considered for a clinical trial if the primary lesion is considered high risk (i.e., stages IIB and IIC). Patients with a positive sentinel lymph node should undergo imaging (CT or PET scanning) to rule out distant metastatic disease, and if none is found (i.e., stage III), adjuvant therapy on or off a clinical study should be offered (see next section). Complete lymphadenectomy for a positive sentinel lymph node has been shown to improve relapse-free but not overall survival, and therefore, it is no longer offered routinely. This avoids the morbidity of regional node dissection in most patients. However, patients not undergoing immediate completion node dissection should have nodal bed surveillance with physical examination and ultrasound at 4- to 6-month intervals for approximately 3 years to rule out isolated nodal bed progression. Complete node dissection is therefore still offered to patients who cannot comply with follow-up and/or forgo adjuvant therapy.

MANAGEMENT OF REGIONALLY METASTATIC MELANOMA (STAGE III)

Patients with a positive sentinel lymph node, resected regional nodal macrometastases, or resected locoregional disease (e.g., recurrences in the wide excision site, within 2 cm of the site ["satellite metastases"], or >2 cm from the site ["in-transit metastases"]) are all considered as having stage III disease. Even after complete resection of stage III disease, the risk for progression to distant metastases (stage IV) may be high, and adjuvant systemic therapy should be offered. Melanomas may recur at the edge of the incision or graft, as satellite metastases, in-transit metastases, or most commonly, regional spread to a draining lymph node basin. Each of these presentations is managed surgically and, increasingly, with post-surgical adjuvant immunotherapy or targeted therapy, after which there is the possibility of long-term disease-free survival. Topical therapy with imiquimod has been useful for patients with low-volume dermal lesions. Talimogene laherparepvec is an engineered, oncolytic herpes simplex virus type 1 that is FDA approved for injection of primary or recurrent melanomas including cutaneous and subcutaneous lesions or lymph node deposits that cannot be completely removed by surgery.

Stage III patients rendered free of disease after surgery are at risk for local or distant recurrence and should be offered adjuvant therapy. Radiotherapy can reduce the risk of local recurrence after lymphadenectomy but does not influence overall survival. Patients with large nodes (>3–4 cm), four or more involved lymph nodes, or extranodal spread on microscopic examination should be considered for radiation. Systemic adjuvant therapy is indicated primarily for patients with stage III disease, but high-risk, node-negative patients (>4 mm thick or ulcerated lesions) and patients with completely resected stage IV disease also may benefit.

Current options for adjuvant therapy include anti-PD-1 (nivolumab or pembrolizumab) or targeted therapy in melanomas that express a *BRAF* V600 mutation. Both anti-PD-1 and targeted therapy have been shown to confer disease-free and overall survival benefits in stage III and stage IV melanoma (see below for further discussion). Other agents such as ipilimumab and interferon α2b (IFN-α2b) have been used in the adjuvant setting, but due to a higher percentage of immune-mediated side effects in the case of ipilimumab and limited efficacy in the case of interferon, they have been supplanted by better alternatives. Ongoing clinical trials are comparing systemic therapy before surgery (neoadjuvant) with adjuvant treatment, the optimal sequence of immunotherapy and targeted therapies, and the utility of anti-PD-1 in high-risk stage

II melanoma. GEP may help to identify patients with stage II or III melanoma who are at lower risk of recurrence and could avoid the toxicity and expense of adjuvant therapy.

TREATMENT

Metastatic Disease

At diagnosis, 84% patients with melanoma will have stage I or II disease and 4% will present with metastases. Many others will develop metastases after initial therapy for locoregional disease. The probability of recurrence is related to initial stage, ranging from <5% with stage IA to >90% for subsets of patients with stage IIID disease at presentation. Patients with a history of melanoma who develop signs or symptoms suggesting recurrent disease should undergo restaging as described earlier. Distant metastases (stage IV) commonly involve skin and lymph nodes as well as viscera, bone, or the brain. The prognosis is better for patients with skin and subcutaneous metastases (M1a) than for lung (M1b) and worst for those with metastases to bone or other visceral organs (M1c) or brain (M1d). An elevated serum LDH is a poor prognostic factor and places the patient in stage M1c regardless of the metastatic sites. The 15-year survival of patients with melanoma was <10% before 2010; however, the development of targeted therapy and immunotherapy has improved disease-free and overall survival, especially for patients with M1a and M1b disease, such that currently the 15-year survival exceeds 25%. Even patients with M1c disease may have prolonged survival, and those who are progression-free for >2 years after immunotherapy or targeted therapy have a high probability of living >5 years from the onset of metastasis.

FDA-approved agents since 2011 include three immune T-cell checkpoint inhibitors (ipilimumab, nivolumab, and pembrolizumab), combination immunotherapy (ipilimumab plus nivolumab), six oral agents that target the MAP kinase pathway (the BRAF inhibitors vemurafenib, dabrafenib, and encorafenib, and the MEK inhibitors trametinib, cobimetinib, and binimetinib), and the oncolytic virus talimogene laherparepvec (**Table 76-4**).

Local modalities, such as surgery and stereotactic radiosurgery, should be considered for patients with oligometastatic disease because they may experience long-term disease-free survival after metastasectomy or ablative high-dose-per-fraction radiation. Patients with solitary metastases are the best candidates, but local modalities can also be used for patients with metastases at more than one site if a complete resection or treatment of all sites can be achieved with reasonable side effects. Patients rendered free of disease can be considered for adjuvant therapy or a clinical trial because their risk of developing additional metastases remains high. Surgery can also be used as an adjunct to systemic therapy when, for example, a few of many metastatic lesions prove resistant to

TABLE 76-4 Treatment Options for Metastatic Melanoma

Immunotherapy
Immune checkpoint blockade
Anti-PD-1: pembrolizumab or nivolumab
Anti-CTLA-4: ipilimumab
Combined ipilimumab and nivolumab
Cytokine-based immunotherapy
High-dose interleukin 2
Oncolytic virus
Talimogene laherparepvec
Targeted therapies
BRAF inhibitors: vemurafenib, dabrafenib, encorafenib
MEK inhibitors: trametinib, cobimetinib, binimetinib
Local modalities
Surgery
Stereotactic radiation

immunotherapy; it may also be helpful to obtain tumor to establish the mutational profile of the recurrent melanoma.

IMMUNOTHERAPY

Checkpoint Blockade Immunotherapies are based on an understanding of the control mechanisms of the normal immune response. Inhibitory receptors or checkpoints, including CTLA-4 and PD-1, are upregulated on T cells after engagement of the T-cell receptor by cognate tumor antigen in the context of the appropriate class I or II HLA molecules during the interaction between a T cell and antigen-presenting cell. Immune checkpoints are an absolute requirement to ensure proper regulation of a normal immune response; however, the continued expression of inhibitory receptors during chronic infection (hepatitis, HIV) and in cancer patients leads to exhausted T cells with limited potential for proliferation, cytokine production, or cytotoxicity. Checkpoint blockade with an antagonistic monoclonal antibody results in improved T-cell function and eradication of tumor cells in preclinical animal models. Ipilimumab, a fully human IgG1 antibody that binds CTLA-4 and blocks inhibitory signals, was the first drug shown in a randomized trial to improve survival in patients with metastatic melanoma. Although response rates are low (about 10%), overall survival is improved. Anti-CTLA-4 monotherapy has been supplanted by combination anti-CTLA-4 plus anti-PD-1 or anti-PD-1 monotherapy due to enhanced survival and, in the case of anti-PD-1 monotherapy, better patient tolerance, as detailed below.

Chronic T-cell activation also leads to induction of PD-1 on the surface of T cells. Expression of one of its ligands, PD-L1, on tumor cells can protect them from immune destruction. Blockade of the PD-1:PD-L1 axis by intravenous (IV) administration of anti-PD-1 or anti-PD-L1 has substantial clinical activity, including cure, in some patients with advanced melanoma and other solid tumors with significantly less toxicity than ipilimumab. The PD-1 blockers, nivolumab and pembrolizumab, have been approved to treat patients with advanced melanoma. Combination T-cell checkpoint therapy, blocking both inhibitory pathways with ipilimumab and nivolumab, leads to superior antitumor activity compared to treatment with either agent alone. Combined therapy with IV ipilimumab and nivolumab is administered in the outpatient setting every 3 weeks for four doses (induction), followed by nivolumab given every 2–4 weeks (maintenance) for up to 1 year, and is associated with an objective response rate of 56% and enhanced survival compared to ipilimumab monotherapy. There may be subsets of patients, specifically those who have >5% expression of PD-1 on T cells in a melanoma biopsy sample, who derive a similar level of clinical benefit from nivolumab monotherapy, although using PD-1 expression to select therapy remains problematic as some patients whose melanoma has no detectable PD-1 expression can still respond to immunotherapy.

T-cell checkpoint antibodies can also interfere with normal immune regulatory mechanisms, which may produce a novel spectrum of side effects. The most common immune-related adverse events were skin rash and diarrhea (sometimes severe, life-threatening colitis), but toxicity can involve almost any organ (e.g., thyroiditis, hypophysitis, hepatitis, nephritis, pneumonitis, myocarditis, neuritis). The severity and frequency of toxicity are greatest with combination T-cell checkpoint antibody therapy, followed by anti-CTLA-4 and then anti-PD-1 monotherapies. Vigilance, interruption of therapy, and early intervention with steroids or other immunosuppressive agents, such as anti-tumor necrosis factor antibodies or mycophenolate mofetil, can mitigate toxicity and prevent permanent organ damage. The management of drug-induced toxicity with immunosuppressive agents does not appear to interfere with antitumor activity, and benefit is manifest even in patients who have to discontinue immunotherapy due to immune-mediated toxicity. The use of T-cell checkpoint antibodies for metastatic melanoma has become commonplace, but there is controversy about whether all patients need combined anti-CTLA-4 and anti-PD-1, whether biomarkers can be used to select patients who may benefit

from anti-PD-1 alone, and the best sequence of targeted therapy and immunotherapy in patients whose melanomas have a *BRAF* mutation. There is also a significant economic impact with any anticancer therapy, which must be placed in the context of the survival benefit.

TARGETED THERAPY

The RAS-RAF-MEK-ERK pathway delivers proliferation and survival signals from the cell surface to the cytoplasm and nucleus and is mutated in approximately 50% of melanomas. Inhibitors of *BRAF* and MEK can induce regression of melanomas that harbor a *BRAF* mutation. Three *BRAF* inhibitors, vemurafenib, dabrafenib, and encorafenib, have been approved for the treatment of patients whose stage IV melanomas harbor a mutation at position 600 in *BRAF*. Monotherapy with *BRAF* inhibitors has been supplanted with combined *BRAF* and MEK inhibition to address the rapid adaptation of the majority of melanomas that use MAP kinase pathway reactivation to facilitate growth when *BRAF* is inhibited. Combined therapy with *BRAF* and MEK inhibitors (dabrafenib and trametinib, vemurafenib with cobimetinib, or encorafenib and binimetinib) improved progression-free and overall survival compared to monotherapy with a *BRAF* inhibitor. Long-term results of inhibition of the MAP kinase pathway confirm that some patients achieve long intervals of disease control, yet the major limitation of both monotherapy and combined therapy appears to be the acquisition of resistance; the majority of patients relapse and eventually die. The mechanisms of resistance are diverse and reflect the genomic heterogeneity of melanoma; however, most instances involve reactivation of the MAPK pathway, often through *RAS* mutations or mutant *BRAF* amplification. Patients who develop resistance to *BRAF* and MEK inhibition are candidates for immunotherapy or clinical trials.

Targeted therapy is accompanied by manageable side effects that differ from those experienced during immunotherapy or chemotherapy. A class-specific side effect of *BRAF* inhibitor monotherapy is the development of hyperproliferative skin lesions, some of which are well-differentiated squamous cell skin cancers (SCCs) occurring in up to 25% of patients. Paradoxical activation of the MAP kinase pathway occurs from *BRAF* inhibitor-mediated changes in *BRAF* wild-type cells, and the activation is blocked by MEK inhibitor, which explains why these lesions occur much less frequently during combined therapy. Metastases of the treatment-induced SCCs have not been reported, and *BRAF* and MEK inhibitors can be continued safely following simple excision of the SCCs. Cardiac and ocular toxicities, although infrequent, can occur with *BRAF* and MEK inhibitors and require medical evaluation, management, and usually discontinuation of targeted therapy.

Activating mutations in the c-kit receptor tyrosine kinase are found in a minority of cutaneous melanomas with chronic sun damage but are more common in mucosal and acral lentiginous subtypes. Overall, the number of patients with *c-kit* mutations is small, but when present, they are similar to those found in gastrointestinal stromal tumors and melanomas with activating *c-kit* mutations and can have clinically meaningful responses to imatinib. The probability of objective response in patients whose melanomas harbor a *c-kit* mutation is 29%. *N-RAS* mutations occur in 15–20% of melanomas. At present, there are no effective targeted agents for these patients, but *N-RAS* inhibitors are being investigated in clinical trials.

Other systemic therapies used to treat stage IV melanoma patients include high-dose interleukin 2, which is also associated with durable remissions in some patients. Chemotherapy with dacarbazine or taxanes is infrequently used, and clinical trials remain an important option for patients with advanced melanoma.

INITIAL APPROACH TO PATIENT WITH METASTATIC DISEASE

Upon diagnosis of stage IV disease, a sample of the patient's tumor should be submitted for molecular testing to determine whether a *BRAF* or *c-kit* mutation is present. Analysis of a metastatic lesion biopsy (if possible) is preferred, but any sample will suffice because

there is little discordance between primary and metastatic lesions. Treatment algorithms start with the tumor's *BRAF* status. For *BRAF* wild-type tumors, immunotherapy is recommended. For patients whose tumors harbor a *BRAF* mutation, initial therapy with either combination *BRAF* and MEK inhibitors or immunotherapy is acceptable. Combined therapy with *BRAF* and MEK inhibitors is recommended for patients with rapidly growing and symptomatic disease when a *BRAF* mutation is present. The sequence of immunotherapy and targeted therapy that confers the greatest survival benefit in patients with minimally symptomatic melanoma is not yet known, but ongoing randomized phase III trials should answer this important question. Despite improvements in therapy, the majority of patients with metastatic melanoma will not be cured, so enrollment in a clinical trial is always an important consideration, even for previously untreated patients.

Clinical trials should be considered for patients with stage IV disease who experience tumor progression despite current therapy. Many will be poor candidates for therapy because of extensive disease burden, poor performance status, or concomitant illness; thus, the timely integration of palliative care and hospice remains an important element of care.

FOLLOW-UP

Skin examination and surveillance at least once a year are recommended for all patients with melanoma. Routine blood work and imaging for patients with stage IA-IIA disease is not recommended unless symptoms are present. Surveillance diagnostic imaging can be considered in patients with stage III high-risk disease but is mainly reserved for patients with signs or symptoms of recurrent disease or to follow response to therapy. For stage-specific recommendations, please consult the NCCN guidelines (see "Further Reading").

NONMELANOMA SKIN CANCERS

NMSCs (mostly SCCs and basal cell cancer [BCC]) are the most common cancers in the United States. Although tumor registries do not routinely gather data on the incidence of NMSCs, it is estimated that the annual incidence is more than 5.3 million cases in the United States; SCCs and BCCs account for 80% and 18%, respectively. While less common, the incidence of Merkel cell carcinoma (MCC) has tripled over the past 20 years. There are now an estimated 1600 cases per year with an annual increase in incidence of 8%. While all forms of NMSCs can metastasize, MCCs do this most commonly, with sentinel lymph node positivity rates of 25% (compared to 12–19% for melanoma) and mortality rates approaching 33% at 3 years. SCCs, particularly those with high-risk features, can also metastasize and account for 2400 deaths annually.

PATHOPHYSIOLOGY AND ETIOLOGY

Similar to melanoma, the most significant cause of NMSCs is UVR, with a dose-response relationship between tanning bed use and the incidence of NMSC. As few as four tanning bed visits per year confers a 15% increase in BCC and an 11% increase in SCC. The risk of lip or oral SCC is increased with cigarette smoking and, like SCC of the ear, has a worse prognosis than that seen on other body sites. Human papillomaviruses and UVR may act as co-carcinogens. Inherited disorders of DNA repair, such as xeroderma pigmentosum, are associated with a greatly increased incidence of skin cancer and help to establish the link between UV-induced DNA damage, inadequate DNA repair, and skin cancer.

The genes damaged most commonly by UV in SCC include *p53* and *N-RAS*, whereas BCC is associated with damage to hedgehog signaling pathway (Hh) genes, which lead to basal cell proliferation. This is usually the result of loss of function of the tumor-suppressor patched homolog 1 (*PTCH1*), which normally inhibits the signaling of smoothened homolog (*SMO*). Two oral *SMO* inhibitors, vismodegib and sonidegib, have been approved by the FDA to treat advanced inoperable or metastatic BCC and locally advanced BCC that has recurred following surgery or radiotherapy, respectively. Vismodegib

also reduces the incidence of BCC in patients with basal cell nevus syndrome who have *PTCH1* mutations, affirming the importance of Hh in the onset of BCC.

Immunosuppression has also been associated with the development of NMSCs; chronically immunosuppressed solid organ transplant recipients have a 65-fold increase in SCC and a 10-fold increase in BCC. The frequency of skin cancer is proportional to the level and duration of immunosuppression and the extent of sun exposure before and after transplantation. SCCs in this population are particularly aggressive, demonstrating higher rates of local recurrence, metastasis, and mortality. Tumor necrosis factor (TNF) antagonist therapy of inflammatory bowel disease and autoimmune disorders, such as rheumatoid and psoriatic arthritis, may also confer an increased risk of NMSC.

Other risk factors for NMSCs include HIV infection, ionizing radiation, thermal burn scars, *BRAF* inhibitor monotherapy, and chronic ulcerations. Albinism, xeroderma pigmentosum, Muir-Torre syndrome, Rombo's syndrome, Bazex-Dupré-Christol syndrome, dyskeratosis congenita, and basal cell nevus syndrome (Gorlin syndrome) also increase the incidence of NMSC.

Although MCC is also clearly related to UV exposure, age, and immunosuppression, this neural crest-derived cancer also appears to have a viral etiology; an oncogenic Merkel cell polyomavirus (MCPyV) is present in 80% of tumors. In patients with MCPyV-positive tumors, there is inactivation of tumor-suppressor genes, specifically the *p53* transcription factor and retinoblastoma protein (*Rb*). In addition, the viral large T antigen is expressed on tumor cells, and many patients have detectable cellular or humoral immune responses to polyoma viral proteins, although this immune response is insufficient to eradicate the malignancy.

CLINICAL PRESENTATION

Basal Cell Carcinoma BCC arises from epidermal basal cells. The least invasive of BCC subtypes, superficial BCC, consists of often subtle, erythematous scaling plaques that slowly enlarge and are most commonly seen on the trunk and proximal extremities (Fig. 76-4). This subtype may be confused with benign inflammatory dermatoses, especially nummular eczema and psoriasis or premalignant actinic keratoses. BCC also can present as a small, slowly growing, pearly nodule, often with tortuous telangiectatic vessels on its surface, rolled borders, and a central crust (nodular BCC). The occasional presence of melanin in this variant of nodular BCC (pigmented BCC) may lead to confusion with melanoma. Morpheaform (fibrosing), infiltrative, and micronodular BCC, the most invasive and potentially aggressive subtypes, manifest as solitary, flat or slightly depressed, indurated whitish, yellowish, or pink scar-like plaques. Borders are typically indistinct, and lesions can be subtle; thus, delay in treatment is common, and tumors can be more extensive than expected clinically. An archaic name for this tumor is "rodent ulcer."

Squamous Cell Carcinoma Primary cutaneous SCC is a malignant neoplasm of keratinizing epidermal cells that has a variable clinical course, ranging from indolent to rapid growth, with the potential to metastasize to regional and distant sites. Commonly, SCC appears as an ulcerated erythematous nodule or superficial erosion on sun-exposed skin of the head, neck, trunk, and extremities (Fig. 76-5). It may also appear as a banal, firm, dome-shaped papule or rough textured plaque. It is commonly mistaken for a wart or callous when the inflammatory response to the lesion is minimal. Dotted or coiled vessels are a hallmark of SCC when viewed through a dermatoscope. The margins of this tumor may be ill defined, and fixation to underlying structures may occur ("tethering").

A very rapidly growing low-grade form of SCC, called keratoacanthoma (KA), typically appears as a large dome-shaped papule with a central keratotic crater. Some KAs regress spontaneously without therapy, but because progression to metastatic SCC has been documented, KAs should be treated in the same manner as other types of cutaneous SCC. KAs occur in 15–25% of patients receiving monotherapy with a *BRAF* inhibitor.

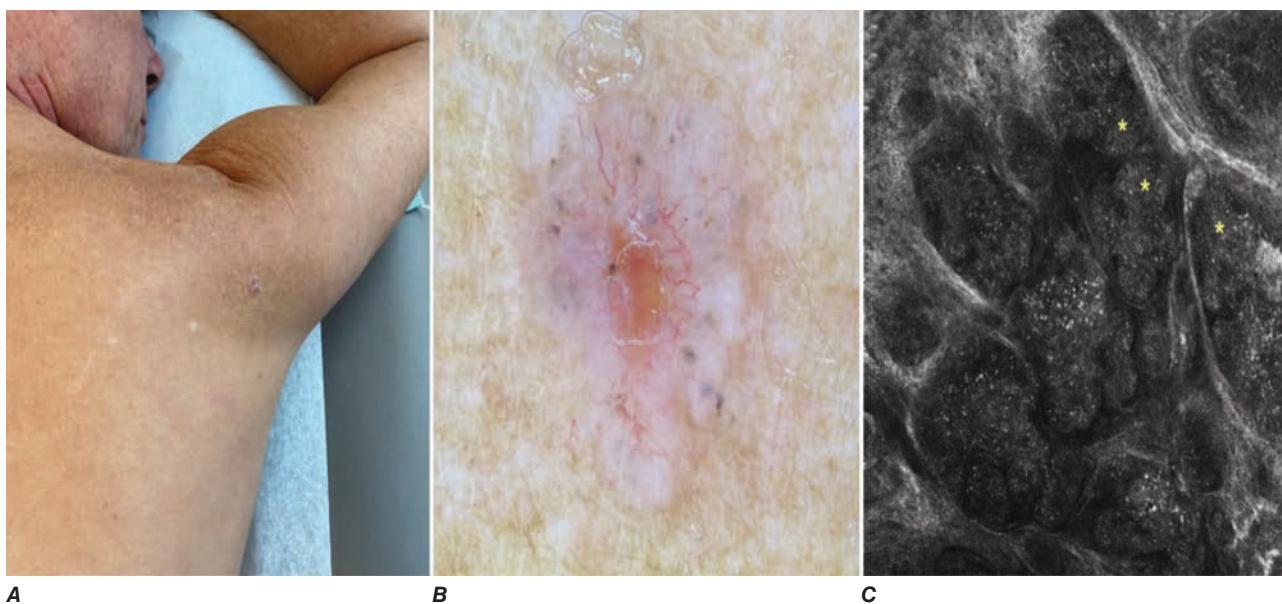


FIGURE 76-4 Clinical and confocal diagnostic findings of basal cell carcinoma. **A.** Typical basal cell carcinoma with skin-colored, slightly translucent rolled borders and a small central erosion on chronically sun-damaged skin of the lateral posterior shoulder. **B.** Dermoscopic image of the same lesion as in panel **A** clearly revealing the central erosion and classic gray, nonreticular globular structures of melanophages that characterize BCC. **C.** In vivo reflectance confocal microscopy of the same lesion as in panel **A** showing typical nests of dermal basaloid cells (*) with classic cleft formation around the nests.

Actinic keratoses and cheilitis (actinic keratoses on the lip), both premalignant forms of SCC, present as hyperkeratotic papules on sun-exposed areas. Malignant transformation occurs in 0.25–20% of untreated lesions. SCC in situ, also called *Bowen's disease*, is the intraepidermal form of SCC and usually presents as a scaling, erythematous plaque. SCC in situ most commonly arises on sun-damaged skin but can occur anywhere on the body. Bowen's disease occurring secondary to infection with human papillomavirus can arise on skin

with minimal or no prior sun exposure, such as the buttock or posterior thigh. Treatment of premalignant and in situ lesions reduces the subsequent risk of invasive disease.

Merkel Cell Carcinoma MCC, also known as cutaneous apudoma, primary neuroendocrine carcinoma of the skin, primary small cell carcinoma of the skin, and trabecular carcinoma of the skin, arises from Merkel cells, which are neuroendocrine skin cells that act as



FIGURE 76-5 Progression of squamous cell carcinoma (SCC). **A.** Actinic keratoses (AKs). **B.** Actinic cheilitis (AK of the lip). **C.** Bowen's disease (SCC in situ). **D.** Keratoacanthoma (well-differentiated SCC). **E.** SCC. **F.** Metastatic SCC.

pressure receptors. Like other skin cancers, MCCs most commonly arise as visible skin lesions, usually as raised, flesh-colored nodules or masses; they can also be red or blue in color and vary in size from 0.5 to >5 cm in diameter and may enlarge rapidly. Although MCCs may arise almost anywhere on the body, they are most commonly found in sun-exposed areas such as the head, neck, or extremities. They can also be found around the anus and on eyelids. The common clinical features of MCC can be summarized by the acronym AEIOU: asymptomatic/nontender, expand rapidly, immune suppression, older than 50 years, and ultraviolet-exposed site.

NATURAL HISTORY

Basal Cell Carcinoma The natural history of BCC is that of a slowly enlarging, locally invasive neoplasm. The degree of local destruction and risk of recurrence vary with the size, duration, location, and histologic subtype of the tumor. Location on the central face, ears, or scalp may portend a higher risk. Small nodular, pigmented, cystic, or superficial BCCs respond well to most treatments. Large lesions and micronodular, infiltrative, and morphaform subtypes may be more aggressive. The metastatic potential of BCC is low (0.1%) in immunocompetent patients, but the risk of recurrence or a new primary NMSC is about 40% over 5 years.

Squamous Cell Carcinoma The natural history of SCC depends on tumor and host characteristics. Tumors arising on sun-damaged skin have a lower metastatic potential than do those on non-sun-exposed areas. Cutaneous SCC metastasizes in 0.3–5.2% of individuals, most frequently to regional lymph nodes. Tumors occurring on the lower lip and ear develop regional metastases in 13 and 11% of patients, respectively, whereas the metastatic potential of SCC arising in scars, chronic ulcerations, and genital or mucosal surfaces is higher. Recurrent SCC has a 30% probability for metastatic spread. Large, poorly differentiated, deep tumors with perineural or lymphatic invasion, multifocal tumors, and those arising in immunosuppressed patients often behave aggressively.

Merkel Cell Carcinoma MCCs have clinical features of both skin cancers and neuroendocrine tumors (particularly small cell lung cancer [SCLC]); thus, they can present locally and develop spread to lymph nodes and distant sites. Molecular markers of neuroendocrine origin such as synaptophysin or chromogranin A are useful to diagnose MCC. Unlike other neuroendocrine tumors, MCCs are not associated with measurable hormone secretion or endocrine syndromes.

Survival with MCC depends on extent of disease: 90% of patients with local disease are cured, whereas 52% with nodal involvement and 10% with distant disease survive. MCC has its own tumor-node-metastasis (TNM) staging system, which incorporates tumor size (<2 cm vs. >2 cm), nodal status (which can be determined by SLNB for clinically negative nodes), and the presence of distant metastases.

Independent of stage, the prognosis of MCC is improved if the tumor cells contain virus, RB protein expression, and intratumoral CD8+ T lymphocyte infiltration, whereas p63 expression, lymphovascular infiltrative pattern, and the presence of immunosuppression (e.g., organ transplant, HIV infection, certain cancers) portend a worse prognosis.

TREATMENT

Basal Cell, Squamous Cell, and Merkle Cell Carcinoma

BASAL CELL CARCINOMA

Treatment for BCC includes electrodesiccation and curettage (ED&C), excision, cryosurgery, radiation therapy (RT), laser therapy, Mohs micrographic surgery (MMS), topical 5-fluorouracil, photodynamic therapy (PDT), and topical immunomodulators, such as imiquimod. The choice of therapy depends on tumor characteristics including depth and location, patient age, medical status, and patient preference. ED&C remains the most commonly employed treatment for superficial, minimally invasive nodular

BCCs and low-risk tumors (e.g., a small tumor of a less aggressive subtype in a favorable location). Wide local excision with standard margins is usually selected for invasive, ill-defined, and more aggressive subtypes of tumors or for cosmetic reasons. MMS, a specialized type of surgical excision that provides the best method for tumor removal while preserving uninvolved tissue, is associated with cure rates >98%. It is the preferred modality for lesions that are recurrent, in high-risk or cosmetically sensitive locations (including recurrent tumors in these locations), and for which maximal tissue conservation is critical (e.g., the eyelids, lips, ears, nose, and digits). RT can cure patients not considered surgical candidates and can be used as a surgical adjunct in high-risk tumors. Imiquimod can be used to treat superficial and smaller nodular BCCs, although it is not FDA approved for nodular BCC. Topical 5-fluorouracil therapy should be limited to superficial BCC. PDT, which uses selective activation of a photoactive drug by visible light, has been used in patients with numerous tumors. Intralesional therapy (5-fluorouracil or IFN) can also be employed. Like RT, it remains an option for selected patients who cannot or will not undergo surgery. Systemic therapy with an SMO inhibitor, vismodegib or sonidegib, is indicated for patients with metastatic or advanced BCC that has recurred after local therapy and who are not candidates for surgery or RT. Targeted therapy with SMO antagonists does not cure patients with BCC but induces regression in approximately 50% of patients with a median duration of response >9 months.

SQUAMOUS CELL CARCINOMA

The principles for surgical management of SCC are the same as for BCC. Previously, advanced disease was treated with cisplatin-containing chemotherapy, intralesional 5-fluorouracil, or cetuximab. These regimens have been supplanted by cemiplimab, a monoclonal antibody targeting PD-1, which causes tumor regression in 47% of patients with advanced disease. SCC and KAs that develop in patients receiving BRAF-targeted therapy should be excised, after which BRAF therapy can be continued.

MERKEL CELL CARCINOMA

The epidemiology, clinical features, and treatments for MCC overlap those for melanoma and NMSC. Early-stage MCCs may be cured with wide local excision of the primary tumor and nodal staging with SLNB. Like SCLCs, MCC is sensitive to radiation, PD-1-directed immunotherapy, and platinum-based chemotherapy. RT is often used as postoperative adjuvant therapy at both the primary excision and SLNB sites, although its use may be withheld around sensitive areas such as the eyelids and hands and after a negative SLNB. For nonsensitive areas, RT may allow for primary excision margins smaller than the traditionally recommended 2-cm radial margins. Similar to melanoma, completion node dissection is now uncommonly used for a positive sentinel node. Adjuvant RT, close observation, and clinical trials investigating immunotherapy based on anti-PD-1 agents are favored.

For patients with metastatic disease, immunotherapy has supplanted chemotherapy. Avelumab (anti-PD-L1) therapy led to objective responses in 33% of patients with advanced MCC; 82% of the responses were durable.

Follow-up of patients with MCC is based on stage and risk. Routine skin exams by a dermatologist familiar with MCC and regular examinations of the nodal basins are recommended. A serum titer of monoclonal antibody to MCPyV should be obtained in newly diagnosed MCC patients. The test can be used to follow patients for relapse if the titer is elevated at baseline and returns to normal after treatment. Conversely, if the titer is elevated but does not return to normal after treatment, imaging should be obtained to look for occult metastases.

PREVENTION

The principles for prevention are those described for melanoma earlier. Unique strategies for NMSC include active surveillance for patients on immunosuppressive medications or BRAF-targeted therapy.



FIGURE 76-6 Other malignant cutaneous tumors. **A.** Patch stage mycosis fungoides (variant of cutaneous T-cell lymphoma). **B.** Tumor stage mycosis fungoides. **C.** Extramammary Paget's disease. **D.** Merkel cell carcinoma. **E.** Dermatofibrosarcoma protuberans. **F.** Kaposi's sarcoma. **G.** Kaposi's sarcoma.

Chemoprophylaxis using synthetic retinoids and immunosuppression reduction when possible may be useful in controlling new lesions and managing patients with multiple tumors. Field therapy with topical 5-fluorouracil, ingenol mebutate, or imiquimod can reduce transformation to SCC in patients with severely sun-damaged skin and numerous premalignant actinic keratoses. Older, immunosuppressed patients should be managed with the lowest doses of immunosuppression possible and encouraged to be particularly careful to minimize UV exposure. Earlier biopsy of unusual-appearing skin lesions may lead to better control of aggressive lesions.

■ OTHER NONMELANOMA CUTANEOUS MALIGNANCIES

Neoplasms of cutaneous adnexae and sarcomas of fibrous, mesenchymal, fatty, and vascular tissues make up the remaining 1–2% of NMSCs (Fig. 76-6). Lymphomas of B- or T-cell origin can also manifest in the skin and can mimic benign conditions such as psoriasis and eczema.

Extramammary Paget's disease is an uncommon apocrine malignancy arising from stem cells of the epidermis that is characterized histologically by the presence of Paget cells. These tumors present as moist erythematous patches on anogenital or axillary skin of the elderly.

Outcomes are generally good with surgery, and 5-year disease-specific survival is 95% with localized disease. Advanced age and extensive disease at presentation confer poorer prognosis. RT or topical imiquimod can be considered for more extensive disease. Local management may be challenging because these tumors often extend far beyond clinical margins; surgical excision with MMS has the highest cure rates. Similarly, MMS is the treatment of choice in other rare cutaneous tumors with extensive subclinical extension such as *dermatofibrosarcoma protuberans*.

Kaposi's sarcoma (KS) is a soft tissue sarcoma of vascular origin that is induced by the human herpesvirus 8. The incidence of KS increased

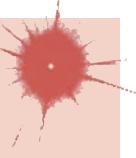
dramatically during the AIDS epidemic, but has now decreased tenfold with the institution of highly active antiretroviral therapy.

ACKNOWLEDGMENT

Walter Urba, MD, PhD, provided valued feedback and suggested improvements to this chapter. Clinical photos were generously provided from the OHSU Swinyer Collection (Leonard Swinyer, MD) and by Drs. Elizabeth Berry, Alexander Witkowski, Joanna Ludzik, Debbie Miller, Alison Skalet, and Justin Leitenberger. Dermoscopic images were provided by Elizabeth Berry, Alexander Witkowski, Joanna Ludzik, and Debbie Miller. Reflectance confocal microscopy images were provided by Drs. Alexander Witkowski and Joanna Ludzik.

■ FURTHER READING

- FARIES MD et al: Completion dissection or observation for sentinel-node metastasis in melanoma. *N Engl J Med* 376:2211, 2017.
- HARMS PW et al: The biology and treatment of Merkel cell carcinoma: Current understanding and research priorities. *Nat Rev Clin Oncol* 15:763, 2018.
- LARKIN J et al: Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 373:23, 2015.
- NATIONAL COMPREHENSIVE CANCER NETWORK: NCCN clinical practice guidelines in oncology (NCCN guidelines): Melanoma. Available from https://www.nccn.org/professionals/physician_gls/pdf/melanoma.pdf. Accessed May 29, 2020.
- ROBERT C et al: Improved overall survival in melanoma with combined dabrafenib and trametinib. *N Engl J Med* 372:30, 2015.
- SHAIN AH, BASTIAN BC: From melanocytes to melanomas. *Nat Rev Cancer* 16:345, 2016.
- WU YP et al: A systematic review of interventions to improve adherence to melanoma preventive behaviors for individuals at elevated risk. *Prev Med* 88:153, 2016.



Epithelial carcinomas of the head and neck arise from the mucosal surfaces in the head and neck and typically are squamous cell in origin. This category includes tumors of the paranasal sinuses, the oral cavity, and the nasopharynx, oropharynx, hypopharynx, and larynx. Tumors of the salivary glands differ from the more common carcinomas of the head and neck in etiology, histopathology, clinical presentation, and therapy. They are rare and histologically highly heterogeneous. Thyroid malignancies are described in [Chap. 385](#).

INCIDENCE AND EPIDEMIOLOGY

The number of new cases of head and neck cancers (oral cavity, pharynx, and larynx) in the United States was estimated at 65,630 in 2020, accounting for about 4% of adult malignancies; estimated deaths were 14,500. The worldwide incidence exceeds half a million cases annually. In North America and Europe, the tumors usually arise from the oral cavity, oropharynx, or larynx. The incidence of oropharyngeal cancers is increasing in recent years, especially in Western countries. Nasopharyngeal cancer is more commonly seen in the Mediterranean countries and in the Far East, where it is endemic in some areas.

ETOLOGY AND GENETICS

Alcohol and tobacco use are the most significant environmental risk factors for head and neck cancer, and when used together, they act synergistically. Smokeless tobacco is an etiologic agent for oral cancers. Other potential carcinogens include marijuana and occupational exposures such as nickel refining, exposure to textile fibers, and woodworking.

Some head and neck cancers have a viral etiology. Epstein-Barr virus (EBV) infection is frequently associated with nasopharyngeal cancer, especially in endemic areas of the Mediterranean and Far East. EBV antibody titers can be measured to screen high-risk populations and are under investigation to monitor treatment response. Nasopharyngeal cancer has also been associated with consumption of salted fish and indoor pollution.

In Western countries, the human papillomavirus (HPV) is associated with a rising incidence of tumors arising from the oropharynx, that is, the tonsillar bed and base of tongue. Over 50% of oropharyngeal tumors are caused by HPV in the United States, and in many urban centers, this proportion is higher. HPV 16 is the dominant viral subtype, although HPV 18 and other oncogenic subtypes are seen as well. Alcohol- and tobacco-related cancers, on the other hand, have decreased in incidence. HPV-related oropharyngeal cancer frequently occurs in a younger patient population and is associated with increased numbers of sexual partners and oral sexual practices. It is associated with a better prognosis, especially for nonsmokers. Vaccination with the nine-valent HPV vaccine may prevent the disease in high-risk populations.

Dietary factors may contribute. The incidence of head and neck cancer is higher in people with the lowest consumption of fruits and vegetables. Certain vitamins, including carotenoids, may be protective if included in a balanced diet. Supplements of retinoids, such as *cis*-retinoic acid, have not been shown to prevent head and neck cancers (or lung cancer) and may increase the risk in active smokers. No specific risk factors or environmental carcinogens have been identified for salivary gland tumors.

HISTOPATHOLOGY, CARCINOGENESIS, AND MOLECULAR BIOLOGY

Squamous cell head and neck cancers are divided into well-differentiated, moderately well-differentiated, and poorly differentiated categories. Poorly differentiated tumors have a worse prognosis than well-differentiated tumors. For nasopharyngeal cancers, the less

common differentiated squamous cell carcinoma is distinguished from nonkeratinizing and undifferentiated carcinoma (lymphoepithelioma) that contains infiltrating lymphocytes and is commonly associated with EBV.

Salivary gland tumors can arise from the major (parotid, submandibular, sublingual) or minor salivary glands (located in the submucosa of the upper aerodigestive tract). Most parotid tumors are benign, but half of submandibular and sublingual gland tumors and most minor salivary gland tumors are malignant. Malignant tumors include mucoepidermoid and adenoid cystic carcinomas and adenocarcinomas.

The mucosal surface of the entire pharynx is exposed to alcohol- and tobacco-related carcinogens and is at risk for the development of a premalignant or malignant lesion. Erythroplakia (a red patch) or leukoplakia (a white patch) can be histopathologically classified as hyperplasia, dysplasia, carcinoma *in situ*, or carcinoma. However, most head and neck cancer patients do not present with a known history of premalignant lesions. Multiple synchronous or metachronous cancers can also be observed. In fact, over time, patients with treated early-stage tobacco- and alcohol-related head and neck cancer are at greater risk of dying from a second malignancy than from a recurrence of the primary disease.

Second head and neck malignancies are usually not therapy induced; they reflect the exposure of the upper aerodigestive mucosa to the same carcinogens that caused the first cancer. These second primaries develop in the head and neck area, the lung, or the esophagus. Thus, computed tomography (CT) screening for lung cancer in heavy smokers who have already developed a head and neck cancer is recommended. Rarely, patients can develop a radiation therapy-induced sarcoma after having undergone prior radiotherapy for a head and neck cancer.

Much progress has been made in describing the molecular features of head and neck cancer. These features have allowed investigators to describe the genetic and epigenetic alterations and the mutational spectrum of these tumors. Early reports demonstrated frequent overexpression of the epidermal growth factor receptor (EGFR). Overexpression was shown to correlate with poor prognosis. However, it has not proved to be a good predictor of tumor response to EGFR inhibitors, which are active in only about 10–15% of patients as single agents. Complex genetic analyses, including those by The Cancer Genome Atlas project, have been performed. *p53* mutations are found frequently with other major affected oncogenic driver pathways including the mitotic signaling and Notch pathways and cell cycle regulation in HPV-negative tumors. HPV oncogenes act through direct inhibition of the *p53* and *RB* tumor-suppressor genes, thereby initiating the carcinogenic process. *HRAS* appears to be emerging as a potentially targetable mutation in a small patient subset. While overall mutation rates are similar in HPV-positive and carcinogen-induced tumors, the specific mutational signature of HPV-positive tumors differs, with frequent alteration of the PI3K pathway and occasional mutations in *KRAS*. Overall, these alterations affect mitogenic signaling, genetic stability, cellular proliferation, and differentiation.

CLINICAL PRESENTATION AND DIFFERENTIAL DIAGNOSIS

Most tobacco-related head and neck cancers occur in patients older than age 60 years. HPV-related malignancies are frequently diagnosed in younger patients, usually in their forties or fifties, whereas EBV-related nasopharyngeal cancer can occur at all ages, including in teenagers. The manifestations vary according to the stage and primary site of the tumor. Patients with nonspecific signs and symptoms in the head and neck area should be evaluated with a thorough otolaryngologic examination, particularly if symptoms persist longer than 2–4 weeks. Males are more frequently affected than women by head and neck cancers, including HPV-positive tumors.

Cancer of the nasopharynx typically does not cause early symptoms. However, it may cause unilateral serous otitis media due to obstruction of the eustachian tube, unilateral or bilateral nasal obstruction, or epistaxis. Advanced nasopharyngeal carcinoma causes neuropathies of the cranial nerves due to skull base involvement.

Carcinomas of the oral cavity present as nonhealing ulcers, changes in the fit of dentures, or painful lesions and masses. Tumors of the tongue base or oropharynx can cause decreased tongue mobility and alterations in speech. Cancers of the oropharynx or hypopharynx rarely cause early symptoms, but they may cause sore throat and/or otalgia. HPV-related tumors frequently present with neck lymphadenopathy as the first sign.

Hoarseness may be an early symptom of laryngeal cancer, and persistent hoarseness requires referral to a specialist for indirect laryngoscopy and/or radiographic studies. If a head and neck lesion treated initially with antibiotics does not resolve in a short period, further workup is indicated; to simply continue the antibiotic treatment may be to lose the chance of early diagnosis of a malignancy.

Advanced head and neck cancers in any location can cause severe pain, otalgia, airway obstruction, cranial neuropathies, trismus, odynophagia, dysphagia, decreased tongue mobility, fistulas, skin involvement, and massive cervical lymphadenopathy, which may be unilateral or bilateral. Some patients have enlarged lymph nodes even though no primary lesion can be detected by endoscopy or biopsy; these patients are considered to have carcinoma of unknown primary (Fig. 77-1). Tonsillectomy and directed biopsies of the base of tongue can frequently identify a small primary tumor that frequently will be HPV related. If the enlarged nodes are located in the upper neck and the tumor cells are of squamous cell histology, the malignancy probably arose from a mucosal surface in the head or neck. Tumor cells in supraclavicular lymph nodes may also arise from a primary site in the chest or abdomen.

The physical examination should include inspection of all visible mucosal surfaces and palpation of the floor of the mouth and of the tongue and neck. In addition to tumors themselves, leukoplakia (a white mucosal patch) or erythroplakia (a red mucosal patch) may be observed; these “ premalignant” lesions can represent hyperplasia, dysplasia, or carcinoma in situ and require biopsy. Further examination should be performed by a specialist. Additional staging procedures include CT or MRI of the head and neck to identify the extent of the disease. Patients with lymph node involvement should have CT scan of the chest and upper abdomen to screen for distant metastases. In heavy smokers, the CT scan of the chest can also serve as a screening tool to rule out a second lung primary tumor. A positron emission tomography (PET) scan can help to identify or exclude distant metastases. CT and PET scans may also be useful in evaluating response to therapy. The definitive staging procedure is an endoscopic examination under anesthesia, which may include laryngoscopy, esophagoscopy,

and bronchoscopy; during this procedure, multiple biopsy samples are obtained to establish a primary diagnosis, define the extent of primary disease, and identify any additional premalignant lesions or second primaries.

Head and neck tumors are classified according to the tumor-node-metastasis (TNM) system of the American Joint Committee on Cancer (AJCC) (Fig. 77-2). This classification varies according to the specific anatomic subsite. In general, primary tumors are classified as T1 to T3 by increasing size, whereas T4 usually represents invasion of another structure such as bone, muscle, or root of tongue. Lymph nodes are staged by size, number, and location (ipsilateral vs contralateral to the primary). Distant metastases are found in <10% of patients at initial diagnosis and are more common in patients with advanced lymph node stage; microscopic involvement of the lungs, bones, or liver is more common, particularly in patients with advanced neck lymph node disease. Modern imaging techniques may increase the number of patients with clinically detectable distant metastases in the future. HPV-related oropharyngeal malignancies have consistently been shown to have a better prognosis, and in the eighth edition of the AJCC staging manual, a separate staging system that takes into account the more favorable outlook of these patients will be included. According to this system, patients with advanced nodal stage can still be considered to have an overall early stage (and associated good prognosis), especially if the patient is a nonsmoker or has limited lifelong tobacco exposure.

In patients with lymph node involvement and no visible primary, the diagnosis should be made by lymph node excision (Fig. 77-1). If the results indicate squamous cell carcinoma, a panendoscopy should be performed, with biopsy of all suspicious-appearing areas and directed biopsies of common primary sites, such as the nasopharynx, tonsil, tongue base, and pyriform sinus. HPV-positive tumors especially can have small primary tumors that spread early to locoregional lymph nodes.

TREATMENT

Head and Neck Cancer

Patients with head and neck cancer can be grossly categorized into three clinical groups: those with localized disease, those with locally or regionally advanced disease (lymph node positive), and those with recurrent and/or metastatic disease below the neck. Comorbidities associated with tobacco and alcohol abuse can affect treatment outcome and define long-term risks for patients who are cured of their disease.

LOCALIZED DISEASE

Nearly one-third of patients have localized disease, that is, T1 or T2 (stage I or stage II) lesions without detectable lymph node involvement or distant metastases. These patients are treated with curative intent by either surgery or radiation therapy. The choice of modality differs according to anatomic location and institutional expertise. Radiation therapy is often preferred for laryngeal cancer to preserve voice function, and surgery is preferred for small lesions in the oral cavity to avoid the long-term complications of radiation, such as xerostomia and osteoradionecrosis and dental decay. Randomized data have shown that a prophylactic staging neck dissection should be part of the surgical procedure to eliminate occult nodal metastatic disease. Overall 5-year survival is 60–90%. Most recurrences occur within the first 2 years following diagnosis and are usually local.

LOCALLY OR REGIONALLY ADVANCED DISEASE

Locally or regionally advanced disease—disease with a large primary tumor and/or lymph node metastases—is the stage of presentation for >50% of patients. Such patients can also be treated with curative intent, but not usually with surgery or radiation therapy alone. Combined-modality therapy, including surgery and/or radiation therapy and chemotherapy, is most successful. Chemotherapy can be administered as induction chemotherapy (chemotherapy before surgery and/or radiotherapy) or as concomitant (simultaneous)

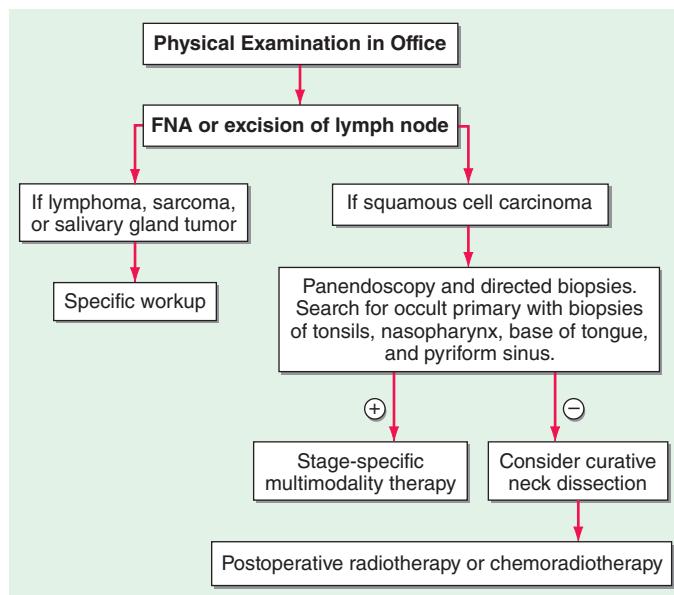


FIGURE 77-1 Evaluation of a patient with cervical adenopathy without a primary mucosal lesion; a diagnostic workup. FNA, fine-needle aspiration.

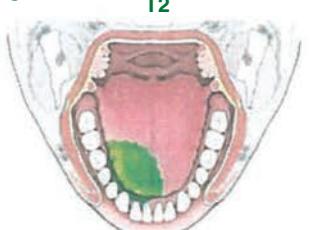
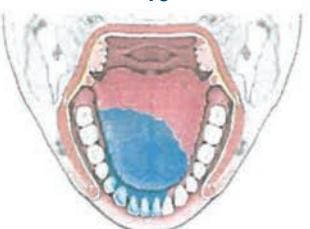
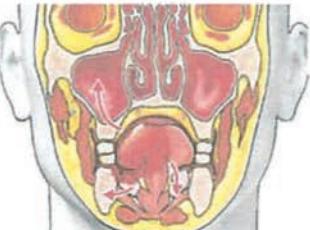
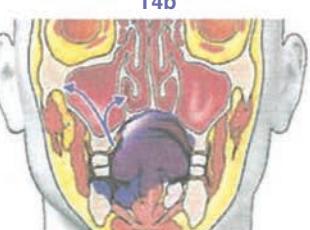
Definition of TNM				Stage groupings		
Stage I  T1	Tumor ≤ 2 cm in greatest dimension ≤ 5 mm depth of invasion (DOI)	 N0	N0- No regional lymph node metastasis	T1	N0	M0
Stage II  T2	Tumor ≥ 2 cm but not more than 4 cm in greatest dimension OR DOI >5 mm and ≤ 10 mm	 N0	N0- No regional lymph node metastasis	T2	N0	M0
Stage III  T3	Tumor ≥ 4 cm OR DOI >10 mm	 N1 ≤ 3 cm	N1- Metastasis in a single ipsilateral lymph node, ≤ 3 cm in greatest dimension	T3	N0	M0
Stage IVA  T4a	Tumor invades skin, mandible, ear canal, fascial nerve, and/or floor of mouth	 N2 ≤ 6 cm	N2a- Metastasis in a single ipsilateral lymph node, >3 cm but ≤ 6 cm N2b- Metastasis in multiple ipsilateral lymph nodes, none >6 cm N2c- Metastasis in bilateral or contralateral lymph nodes, none >6 cm	T4a	N0	M0
Stage IVA  T4b	Tumor invades skull base and/or pterygoid plates and/or encases carotid artery	 N3 >6 cm	N3- Metastasis in a lymph node >6 cm in greatest dimension or clinically overt extranodal extension	T4b	Any N	M0
Stage IVC		M1		Any T	Any N	M1

FIGURE 77-2 Tumor-node-metastasis (TNM) staging system. (Figure based on the AJCC Cancer Staging Manual, 8th edition.)

chemotherapy and radiation therapy. The latter is currently most commonly used and supported by the best evidence. Five-year survival rates exceed 50% in many trials, but part of this increased survival may be due to an increasing fraction of study populations with HPV-related tumors who carry a better prognosis. HPV testing of newly diagnosed tumors should be performed for patients with oropharyngeal tumors at the time of diagnosis. Clinical trials for HPV-related tumors are focused on exploring reductions in

treatment intensity, especially radiation dose, in order to ameliorate long-term toxicities (fibrosis, swallowing dysfunction).

In patients with intermediate-stage tumors (stage III and early stage IV), concomitant chemoradiotherapy can be administered as a primary treatment for patients with unresectable disease, to pursue an organ-preserving approach especially for patients with laryngeal cancer (omission of surgery), or in the postoperative setting for smaller resectable tumors with adverse prognostic features.

Induction Chemotherapy In this strategy, patients receive chemotherapy (current standard is a three-drug regimen of docetaxel, cisplatin, and fluorouracil [5-FU]) before surgery and radiation therapy. Most patients who receive three cycles show tumor reduction, and the response is clinically “complete” in up to half of patients. This “sequential” multimodality therapy allows for organ preservation in patients with laryngeal and hypopharyngeal cancer and results in higher cure rates compared with radiotherapy alone.

Concomitant Chemoradiotherapy With the concomitant strategy, chemotherapy and radiation therapy are given simultaneously rather than in sequence. Tumor recurrences from head and neck cancer develop most commonly locoregionally (in the head and neck area of the primary and draining lymph nodes). The concomitant approach is aimed at enhancing tumor cell killing by radiation therapy in the presence of chemotherapy (radiation enhancement) and is a conceptually attractive approach for bulky tumors. Toxicity (especially mucositis, grade 3 or 4, in 70–80%) is increased with concomitant chemoradiotherapy. However, meta-analyses of randomized trials document an improvement in 5-year survival of 8% with concomitant chemotherapy and radiation therapy. Cisplatin is preferentially given weekly during a course of daily radiotherapy over a 6- to 7-week course. In addition, concomitant chemoradiotherapy produces better laryngectomy-free survival (organ preservation) than radiation therapy alone in patients with advanced larynx cancer. The use of radiation therapy together with cisplatin produces improved survival in patients with advanced nasopharyngeal cancer. The outcome of HPV-related cancers seems to be especially favorable following cisplatin-based chemoradiotherapy. Trials substituting cisplatin with the EGFR inhibitor cetuximab in that patient population have shown inferior survival.

The success of concomitant chemoradiotherapy in patients with unresectable disease has led to the testing of a similar approach in patients with resected intermediate-stage disease as a postoperative therapy. Concomitant chemoradiotherapy produces a significant improvement over postoperative radiation therapy alone for patients whose tumors demonstrate higher risk features, such as extracapsular spread beyond involved lymph nodes, involvement of multiple lymph nodes, or positive margins at the primary site following surgery.

A monoclonal antibody to EGFR (cetuximab) increases survival rates when administered during radiotherapy. EGFR blockade results in radiation sensitization and has milder systemic side effects than traditional chemotherapy agents, although an acneiform skin rash is commonly observed. Nevertheless, the addition of cetuximab to current standard chemoradiotherapy regimens has failed to show further improvement in survival and is not recommended.

TREATMENT APPROACHES FOR HPV-RELATED HEAD AND NECK CANCERS

Given consistent observations of high survival rates for patients with advanced HPV-related oropharyngeal tumors using combined-modality treatment strategies, de-escalation protocols have attracted widespread interest. The goal here is to decrease the long-term morbidity resulting from high-dose radiation therapy, including extensive neck fibrosis, swallowing problems, and osteoradionecrosis of the jaw. Current studies are investigating the use of lower radiation doses, the use of induction chemotherapy and subsequent omission of chemotherapy or administration of significantly reduced chemoradiation doses in very good responders, and other strategies. In addition, interest has increased in surgical approaches using robotic surgery, which allows better visualization of the base of tongue and tonsil. While technically feasible, this approach remains investigational because a large number of patients with disease involving multiple lymph nodes will still require postoperative chemoradiotherapy, thus negating the goal of treatment de-escalation. It is expected that distinct treatment guidelines from carcinogen-induced tumors will be defined in the coming years.

RECURRENT AND/OR METASTATIC DISEASE

Five to ten percent of patients present with metastatic disease, and 30–50% of patients with locoregionally advanced disease experience recurrence, frequently outside the head and neck region. Patients with recurrent and/or metastatic disease are, with few exceptions, treated with palliative intent. Some patients may require local or regional radiation therapy for pain control, but most are given systemic therapy.

Combination chemotherapy formerly was the first-line systemic therapeutic approach to patients with recurrent disease after prior curative intent surgery and/or chemoradiotherapy or those presenting initially with metastatic disease. In particular, a combination of cisplatin with 5-FU and cetuximab (the EXTREME regimen) was frequently used.

However, immunotherapies have proven to be of value in this setting. In particular, inhibitors of the immunosuppressive lymphocyte surface receptor (PD-1) pathway have shown activity in squamous cell cancers of the head and neck. A randomized trial evaluating the PD-1 inhibitor nivolumab versus traditional chemotherapy in the second-line treatment of patients with recurrent or metastatic disease showed a significant increase in 1-year survival rates with fewer severe treatment-related toxicities. In addition, some responses were of long duration, allowing a cohort of patients to live far beyond the historical median of <1 year. The PD-1 inhibitor pembrolizumab also demonstrated activity in a similarly designed randomized trial.

Following establishment of second-line activity, pembrolizumab was compared as single-agent therapy or in combination with cisplatin and 5-FU with prior standard chemotherapy alone (cisplatin, 5-FU, and cetuximab). In this trial, overall survival was improved with pembrolizumab versus chemotherapy as well as with the combination of chemotherapy plus pembrolizumab. No statistically significant impact on progression-free survival was noted. In addition, expression of PD-L1 in the tumor tissue was shown to be of importance. Patients with tumors high in expression (PD-L1 score >20%; i.e., expression of PD-L1 on 20% of tumor cells) had a marked survival benefit with pembrolizumab as single agent, whereas patients with lower PD-L1 expression had a less impressive but still statistically significant survival benefit. However, for the group expressing lower levels of PD-L1, the combination of pembrolizumab with chemotherapy showed more substantial benefit. Current standard treatment therefore frequently consists of combination chemoimmunotherapy for patients with a low combined positive score (CPS; the fraction of tumor cells expressing PD-L1), whereas those with higher CPS scores can be treated with immunotherapy alone, especially if overall tumor burden is limited. Patients who experience progression after first-line chemoimmunotherapy or immunotherapy can then be treated with additional single-agent or combination chemotherapy.

EGFR-directed therapies, including monoclonal antibodies (e.g., cetuximab) and tyrosine kinase inhibitors (TKIs) of the EGFR signaling pathway (e.g., erlotinib or gefitinib), have single-agent activity of ~10%. Side effects are usually limited to an acneiform rash and diarrhea (for the TKIs). The addition of cetuximab to standard combination chemotherapy with cisplatin or carboplatin and 5-FU results in a significant increase in median survival. Drugs targeting specific mutations are under investigation, and patients with HRAS mutations have tumor shrinkage with the farnesyltransferase inhibitor tipifarnib.

COMPLICATIONS

Complications from treatment of head and neck cancer are usually correlated to the extent of surgery and exposure of normal tissue structures to radiation. Currently, the extent of surgery has been limited or completely replaced by chemotherapy and radiation therapy as the primary approach. Acute complications of radiation include mucositis and dysphagia. Long-term complications include xerostomia, loss of taste, decreased tongue mobility, second malignancies, dysphagia, and neck fibrosis. The complications of

chemotherapy vary with the regimen used but usually include myelosuppression, mucositis, nausea and vomiting, and nephrotoxicity (with cisplatin).

The mucosal side effects of therapy can lead to malnutrition and dehydration. Many centers address issues of dentition before starting treatment, and some place feeding tubes to ensure control of hydration and nutrition intake. About 50% of patients develop hypothyroidism from the treatment; thus, thyroid function should be monitored.

SALIVARY GLAND TUMORS

Most benign salivary gland tumors are treated with surgical excision, and patients with invasive salivary gland tumors are treated with surgery and radiation therapy. These tumors may recur regionally; adenoid cystic carcinoma has a tendency to recur along the nerve tracks. Distant metastases may occur as late as 10–20 years after the initial diagnosis. For metastatic disease, therapy is given with palliative intent, usually chemotherapy with doxorubicin and/or cisplatin. Identification of novel agents with activity in these tumors is a high priority. It is hoped that comprehensive genomic characterization of these rare tumors will facilitate these efforts.

FURTHER READING

- AGRAWAL N et al: Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* 333:1154, 2011.
- BURTNES B et al: Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): A randomized, open-label, phase 3 study. *Lancet* 394:1915, 2019.
- CHOW LQM: Head and neck cancer. *N Engl J Med* 382:60, 2020.
- D'CRUZ AK et al: Elective versus therapeutic neck dissection in node-negative oral cancer. *N Engl J Med* 373:521, 2015.
- FERRIS RL et al: Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N Engl J Med* 375:1856, 2016.
- GILLISON ML et al: Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *J Natl Cancer Inst* 100:407, 2008.
- KANG H et al: Whole-exome sequencing of salivary gland mucoepidermoid carcinoma. *Clin Cancer Res* 23:283, 2017.
- MEHANNA H et al: De-escalation after DE-ESCALATE and RTOG 1016: A Head and Neck Cancer Intergroup framework for future de-escalation studies. *J Clin Oncol* 38:2552, 2020.
- SABATINI ME et al: Human papillomavirus as a driver of head and neck cancers. *Br J Cancer* 122:306, 2020.
- TOTA JE et al: Evolution of the oropharynx cancer epidemic in the United States: Moderation of increasing incidence in younger individuals and shift in the burden to older individuals. *J Clin Oncol* 37:1538, 2019.

established in the mid-twentieth century and codified with the release of the U.S. Surgeon General's 1964 report on the health effects of tobacco smoking. Following the report, cigarette use started to decline in North America and parts of Europe, and with it, so did the incidence of lung cancer. Although tobacco smoking remains the primary cause of lung cancer worldwide, approximately 60% of new lung cancers in the United States occur in former smokers (smoked ≥100 cigarettes per lifetime, quit ≥1 year), many of whom quit decades ago, or never smokers (smoked <100 cigarettes per lifetime). Moreover, one in five women and one in 12 men diagnosed with lung cancer have never smoked.

EPIDEMIOLOGY

Lung cancer is the most common cause of cancer death among American men and women. Approximately 228,000 individuals will be diagnosed with lung cancer in the United States in 2020, and >135,000 individuals will die from the disease. Lung cancer is uncommon below age 40, with rates increasing until age 80, after which the rate tapers off. The projected lifetime probability of developing lung cancer is estimated to be ~8% among males and ~6% among females. The incidence of lung cancer varies by racial and ethnic group, with the highest age-adjusted incidence rates among African Americans. The excess in age-adjusted rates among African Americans occurs only among men, but examinations of age-specific rates show that below age 50, mortality from lung cancer is >25% higher among African American than Caucasian women. Incidence and mortality rates among Hispanics and Native and Asian Americans are ~40–50% those of whites.

RISK FACTORS

Cigarette smokers have a 10-fold or greater increased risk of developing lung cancer compared to those who have never smoked. A large-scale genomic study suggested that one genetic mutation is induced for every 15 cigarettes smoked. The risk of lung cancer is lower among persons who quit smoking than among those who continue smoking. The size of the lung cancer risk reduction increases with the length of time the person has quit smoking, although even long-term former smokers have higher risks of lung cancer than those who never smoked. Cigarette smoking has been shown to increase the risk of all major types of lung cancer. Environmental tobacco smoke (ETS) or second-hand smoke is also an established cause of lung cancer. The risk from ETS is less than from active smoking, with about a 20–30% increase in lung cancer observed among never smokers married for many years to smokers, in comparison to the 2000% increase among continuing active smokers. The impact on the development of lung cancer among users of alternate nicotine delivery devices (e-cigarettes or vaping) is undefined. While one large randomized study demonstrated the superiority of e-cigarettes compared to traditional nicotine replacement therapy in aiding smoking cessation, e-cigarette- or vaping-associated lung injury (EVALI) is an emerging phenomenon that poses risks that may counterbalance the potential benefit in helping patients reduce traditional cigarette consumption and lung cancer risk.

Although cigarette smoking is the cause of the majority of lung cancers, several other risk factors have been identified, including occupational exposure to asbestos, arsenic, bischloromethyl ether, hexavalent chromium, mustard gas, nickel (as in certain nickel-refining processes), and polycyclic aromatic hydrocarbons.

Ionizing radiation is also an established lung carcinogen, most convincingly demonstrated from studies showing increased rates of lung cancer among survivors of the atom bombs dropped on Hiroshima and Nagasaki and large excesses among workers exposed to alpha irradiation from radon in underground uranium mining. Prolonged exposure to low-level radon in homes might impart a risk of lung cancer equal to or greater than that of ETS. Prior lung diseases such as chronic bronchitis, emphysema, and tuberculosis have been linked to increased risks of lung cancer as well. The risk of lung cancer appears to be higher among individuals with low fruit and vegetable intake during adulthood. This observation led to hypotheses that specific nutrients, in particular retinoids and carotenoids, might have chemopreventative effects for lung cancer. However, randomized trials failed to validate this hypothesis.

78

Neoplasms of the Lung

Leora Horn, Wade T. Iams



Lung cancer, which was rare before 1900 with fewer than 400 cases described in the medical literature, is considered a disease of modern man, killing over three times as many men as prostate cancer and nearly twice as many women as breast cancer. Although lung cancer remains the number one cause of cancer-related mortality, a decline in lung cancer deaths has emerged, attributed to improvements in testing and therapeutic strategies and a decline in tobacco usage. Tobacco consumption is the primary cause of lung cancer, a reality firmly

Smoking Cessation Given the undeniable link between cigarette smoking and lung cancer, physicians must promote tobacco abstinence. Stopping tobacco use before middle age avoids >90% of the lung cancer risk attributable to tobacco. Importantly, smoking cessation can even be beneficial in individuals with an established diagnosis of lung cancer, as it is associated with improved survival, fewer side effects from therapy, and an overall improvement in quality of life. Consequently, it is important to promote smoking cessation even *after* the diagnosis of lung cancer is established.

Physicians need to understand the essential elements of smoking cessation therapy. The individual must want to stop smoking and must be willing to work hard to achieve the goal of smoking abstinence. Self-help strategies alone only marginally affect quit rates, whereas individual and combined pharmacotherapies in combination with counseling can significantly increase rates of cessation. Therapy with an antidepressant (e.g., bupropion) and nicotine replacement therapy (varenicline, a $\alpha_4\beta_2$ nicotinic acetylcholine receptor partial agonist) are approved by the U.S. Food and Drug Administration (FDA) as first-line treatments for nicotine dependence. In a randomized trial, varenicline was shown to be more efficacious than bupropion or placebo. Prolonged use of varenicline beyond the initial induction phase proved useful in maintaining smoking abstinence. Clonidine and nortriptyline are recommended as second-line treatments. A role for e-cigarettes has not been definitively established (**Chap. 454**).

Inherited Predisposition to Lung Cancer Exposure to environmental carcinogens, such as those found in tobacco smoke, induce or facilitate the transformation from bronchoepithelial cells to a malignant phenotype. The contribution of carcinogens to transformation is modulated by polymorphic variations in genes that affect aspects of carcinogen metabolism. Certain genetic polymorphisms of the P450 enzyme system, specifically CYP1A1, and chromosome fragility are associated with the development of lung cancer. These genetic variations occur at relatively high frequency in the population, but their contribution to an individual's lung cancer risk is generally low. However, because of their population frequency, the overall impact on lung cancer risk could be high.

First-degree relatives of lung cancer probands have a two- to three-fold excess risk of lung cancer and other cancers, many of which are not smoking-related. These data suggest that specific genes and/or genetic variants may contribute to susceptibility to lung cancer. However, very few such genes have yet been identified. Individuals with inherited mutations in *RB* (patients with retinoblastoma living to adulthood) and *p53* (Li-Fraumeni syndrome) genes may develop lung cancer. Common gene variants involved in lung cancer have identified three separate loci that are associated with lung cancer (5p15, 6p21, and 15q25) and include genes that regulate acetylcholine nicotinic receptors and telomerase production. A rare germline mutation (T790M) involving the epidermal growth factor receptor (EGFR) maybe be linked to lung cancer susceptibility in never smokers. Likewise, a susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers. Although progress has been made, there is a significant amount of work that remains to be done in identifying heritable risk factors for lung cancer. Currently no molecular criteria are suitable to select patients for more intense screening programs or for specific chemopreventive strategies.

PATHOLOGY

The World Health Organization (WHO) defines lung cancer as tumors arising from the respiratory epithelium (bronchi, bronchioles, and alveoli). The WHO classification system divides epithelial lung cancers into four major cell types: small-cell lung cancer (SCLC), adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma; the latter three types are collectively known as non-small-cell carcinomas (NSCLCs) (**Fig. 78-1**). Small-cell carcinomas consist of small cells with scant cytoplasm,

ill-defined cell borders, finely granular nuclear chromatin, absent or inconspicuous nucleoli, and a high mitotic count. SCLC may be distinguished from NSCLC by the presence of neuroendocrine markers including CD56, neural cell adhesion molecule (NCAM), synaptophysin, and chromogranin. Adenocarcinomas possess glandular differentiation or mucin production and may show acinar, papillary, lepidic, or solid features or a mixture of these patterns. Squamous cell carcinomas of the lung are morphologically identical to extrapulmonary squamous cell carcinomas and cannot be distinguished by immunohistochemistry alone. Squamous cell tumors show keratinization and/or intercellular bridges that arise from bronchial epithelium. The tumor consists of sheets of cells rather than the three-dimensional groups of cells characteristic of adenocarcinomas. Large-cell carcinomas compose <10% of lung carcinomas. These tumors lack the cytologic and architectural features of small-cell carcinoma and glandular or squamous differentiation. Together, these four histologic types account for ~90% of all epithelial lung cancers.

All histologic types of lung cancer can develop in current and former smokers, although squamous and small-cell carcinomas are most commonly associated with tobacco use. With the decline in cigarette consumption, adenocarcinoma has become the most frequent histologic subtype of lung cancer in the United States. In lifetime never smokers or former light smokers (<10 pack-year history), women, and younger adults (<60 years), adenocarcinoma tends to be the most common form of lung cancer.

In addition to distinguishing between SCLC and NSCLC, because these tumors have quite different natural histories and therapeutic approaches (see below), it is necessary to classify whether NSCLC is squamous or nonsquamous. The classification system, developed jointly by the International Association for the Study of Lung Cancer, the American Thoracic Society, and the European Respiratory Society, provides an integrated approach to the classification of lung adenocarcinoma that includes clinical, molecular, radiographic, and pathologic information.

It is recognized that most lung cancers present in an advanced stage and are often diagnosed based on small biopsies or cytologic specimens, rendering clear histologic distinctions difficult, if not impossible. In such cases, particularly in patients with advanced-stage disease, a repeat biopsy is recommended to obtain additional tissue for further clarification. The distinction between squamous and nonsquamous lung cancer is viewed as critical to optimal therapeutic decision making, and a diagnosis of *non-small-cell carcinoma, not otherwise specified* is no longer considered acceptable. This distinction can be achieved using a single marker for adenocarcinoma (thyroid transcription factor-1 or napsin-A) plus a squamous marker (p40 or p63) and/or mucin stains. If tissue is limited and a clear morphologic pattern is evident, a diagnosis can be made without immunohistochemistry staining. In addition to determining histologic subtype, preservation of sufficient specimen material for appropriate molecular testing and PD-L1 testing necessary to help guide therapeutic decision making is recommended (see below).

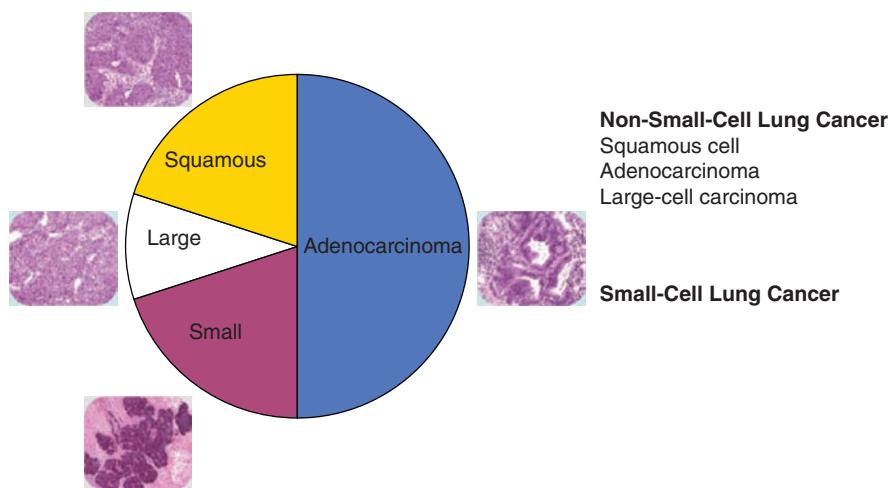


FIGURE 78-1 Histologic subsets of lung cancer.

The terms *adenocarcinoma in situ* and *minimally invasive adenocarcinoma* are now recommended for small solitary adenocarcinomas (≤ 3 cm) with either pure lepidic growth (term used to describe single-layered growth of atypical cuboidal cells coating the alveolar walls) or predominant lepidic growth with ≤ 5 mm invasion. Individuals with these entities experience 100% or near 100% 5-year disease-free survival with complete tumor resection. *Invasive adenocarcinomas*, representing more than 70–90% of surgically resected lung adenocarcinomas, are now classified by their predominant pattern: lepidic, acinar, papillary, and solid patterns. Lepidic-predominant subtype has a favorable prognosis, acinar and papillary have an intermediate prognosis, and solid-predominant has a poor prognosis. The terms *signet ring* and *clear cell adenocarcinoma* have been eliminated from the variants of invasive lung adenocarcinoma, whereas the term *micropapillary*, a subtype with a particularly poor prognosis, has been added. Because of prognostic implications, squamous cell carcinoma has also been modified to consist of keratinizing, nonkeratinizing, and basaloid, analogous to head and neck cancers.

IMMUNOHISTOCHEMISTRY

The diagnosis of lung cancer most often rests on the morphologic or cytologic features correlated with clinical and radiographic findings. Immunohistochemistry may be used to verify neuroendocrine differentiation within a tumor with markers such as neuron-specific enolase (NSE), CD56 or NCAM, synaptophysin, chromogranin, and Leu7. Immunohistochemistry is also helpful in differentiating primary from metastatic adenocarcinomas; thyroid transcription factor-1 (TTF-1), identified in tumors of thyroid and pulmonary origin, is positive in >70% of pulmonary adenocarcinomas and is a reliable indicator of primary lung cancer, provided a thyroid primary has been excluded. A negative TTF-1, however, does not exclude the possibility of a lung primary. TTF-1 is also positive in neuroendocrine tumors of pulmonary and extrapulmonary origin. Napsin-A (Nap-A) is an aspartic protease that plays an important role in maturation of surfactant B7 and is expressed in cytoplasm of type II pneumocytes. In several studies, Nap-A has been reported in >90% of primary lung adenocarcinomas. Notably, a combination of Nap-A and TTF-1 is useful in distinguishing primary lung adenocarcinoma (Nap-A positive, TTF-1 positive) from primary lung squamous cell carcinoma (Nap-A negative, TTF-1 negative) and primary SCLC (Nap-A negative, TTF-1 positive). Cytokeratins 7 and 20 used in combination can help narrow the differential diagnosis; nonsquamous NSCLC, SCLC, and mesothelioma may stain positive for CK7 and negative for CK20, whereas squamous cell lung cancer often will be both CK7 and CK20 negative. p63 is a useful marker for the detection of NSCLCs with squamous differentiation when used in cytologic pulmonary samples. Mesothelioma can be easily identified ultrastructurally, but it has historically been difficult to differentiate from adenocarcinoma through morphology and immunohistochemical staining. Several markers in the past few years have proven to be more helpful including CK5/6, calretinin, and Wilms tumor gene-1 (WT1), all of which show positivity in mesothelioma.

MOLECULAR PATHOGENESIS

As proposed by Hanahan and Weinberg, virtually all cancer cells acquire six hallmark capabilities: self-sufficiency in growth signals, insensitivity to antigrowth signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. The order in which these hallmark capabilities are acquired is variable. Events leading to acquisition of these hallmarks vary widely, although broadly, cancers arise as a result of accumulations of gain-of-function mutations in oncogenes and loss-of-function mutations in tumor-suppressor genes. Further complicating

the study of lung cancer, the sequence of events that leads to disease is clearly different for the various histopathologic entities.

For cancers in general, one theory holds that a small subset of the cells within a tumor (i.e., “stem cells”) are responsible for the full malignant behavior of the tumor. As part of this concept, the large bulk of the cells in a cancer are “offspring” of these cancer stem cells. While clonally related to the cancer stem cell subpopulation, most cells by themselves cannot regenerate the full malignant phenotype. The stem cell concept may explain the failure of standard medical therapies to eradicate lung cancers, even when there is a clinical complete response. Disease recurs because therapies do not eliminate the stem cell component, which may be more resistant to therapy. Precise human lung cancer stem cells have yet to be identified.

Among lung cancer histologies, adenocarcinomas have been the most extensively catalogued for recurrent genomic gains and losses as well as for somatic mutations (Fig. 78-2, Table 78-1). While multiple different kinds of aberrations have been found, a major class involves “driver mutations,” which are mutations that occur in genes encoding signaling proteins that, when aberrant, drive initiation and maintenance of tumor cells. Importantly, driver mutations can serve as a potential Achilles’ heels for tumors, if their gene products can be targeted appropriately. These genes encode cell-surface receptors consisting of an extracellular ligand-binding domain, a transmembrane structure, and an intracellular tyrosine kinase (TK) domain. The binding of ligand to receptor activates receptor dimerization and TK autophosphorylation, initiating a cascade of intracellular events, and leading to increased cell proliferation, angiogenesis, metastasis, and a decrease in apoptosis. Lung adenocarcinomas can arise when normal alveolar type II cells develop mutations in EGFR, BRAF, MET, KRAS, and PIK3CA. These same tumors display high sensitivity to small-molecule TK inhibitors (TKIs). Additional subsets of lung adenocarcinoma have been identified as defined by the presence of specific chromosomal rearrangements, resulting in the aberrant activation of the TKs ALK, ROS1, NTRK, and RET. Notably, most driver mutations in lung cancer appear to be mutually exclusive, suggesting that acquisition of one of these mutations is sufficient to drive tumorigenesis. Although driver mutations have mostly been found in adenocarcinomas, three

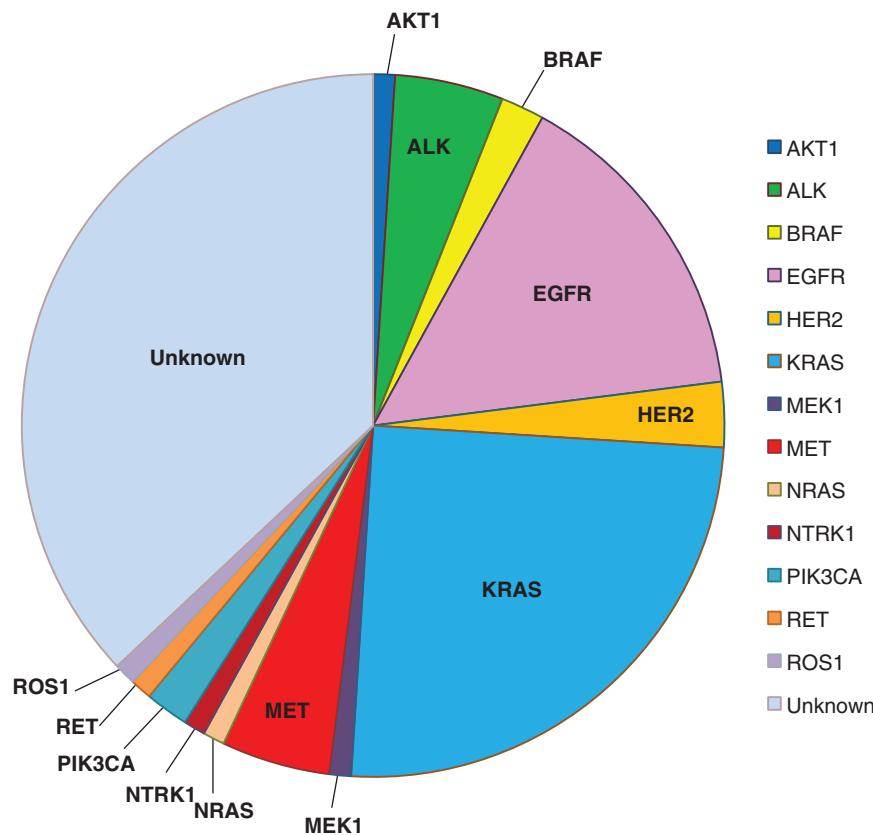


FIGURE 78-2 Driver mutations in lung adenocarcinomas.

TABLE 78-1 Driver Mutations in Non-Small-Cell Lung Cancer (NSCLC)

GENE	ALTERATION	FREQUENCY IN NSCLC	TYPICAL HISTOLOGY
AKT1	Mutation	1%	Adenocarcinoma, squamous
ALK	Rearrangement	3–7%	Adenocarcinoma
BRAF	Mutation	1–3%	Adenocarcinoma
DDR2	Mutation	~4%	Squamous
EGFR	Mutation	10–35%	Adenocarcinoma
FGFR1	Amplification	~20%	Squamous
HER2	Mutation	2–4%	Adenocarcinoma
KRAS	Mutation	15–25%	Adenocarcinoma
MEK1	Mutation	1%	Adenocarcinoma
MET	Amplification	2–4%	Adenocarcinoma
NRAS	Mutation	1%	Adenocarcinoma
NTRK	Rearrangement	1–2%	Adenocarcinoma
PIK3CA	Mutation	1–3%	Squamous
PTEN	Mutation	4–8%	Squamous
ROS1	Rearrangement	1–2%	Adenocarcinoma

potential molecular targets have been identified in squamous cell lung carcinomas: *FGFR1* amplification, *DDR2* mutations, and *PIK3CA* mutations/*PTEN* loss as well as *BRAF* and *MET* (Table 78-1).

A large number of tumor-suppressor genes have also been identified that are inactivated during the pathogenesis of lung cancer. These include *TP53*, *RB1*, *RASSF1A*, *CDKN2A/B*, *LKB1* (*STK11*), and *FHIT*. Nearly 90% of SCLCs harbor mutations in *TP53* and *RB1*. Several tumor-suppressor genes on chromosome 3p appear to be involved in nearly all lung cancers. Allelic loss of this region occurs very early in lung cancer pathogenesis, including in histologically normal smoking-damaged lung epithelium.

EARLY DETECTION AND SCREENING

In lung cancer, clinical outcome is related to the stage at diagnosis, and hence, it is generally assumed that early detection of occult tumors will lead to improved survival. Early detection is a process that involves screening tests, surveillance, diagnosis, and early treatment. Screening refers to the use of tests across a healthy population in order to identify individuals who harbor asymptomatic disease. For a screening program to be successful, the target population must have a high burden of disease; the test must be sensitive, specific, accessible, and cost effective; and effective treatment must be available that can reduce mortality. With any screening procedure, it is important to consider the possible influence of *lead-time bias* (detecting the cancer earlier without an effect on survival), *length-time bias* (indolent cancers are detected on screening and may not affect survival, whereas aggressive cancers are likely to cause symptoms earlier in patients and are less likely to be detected), and *overdiagnosis* (diagnosing cancers so slow growing that they are unlikely to cause the death of the patient).

Because a majority of lung cancer patients present with advanced disease beyond the scope of surgical resection, the value of screening for this condition is debated. Indeed, randomized controlled trials conducted in the 1960s to 1980s using screening chest x-rays (CXR), with or without sputum cytology, reported no impact on lung cancer-specific mortality in patients characterized as high risk (males age ≥ 45 years with a smoking history). These studies have been criticized for their design, statistical analyses, and outdated imaging modalities. In contrast to CXR, low-dose, noncontrast, thin-slice spiral chest computed tomography (LDCT) has emerged as an effective tool to screen for lung cancer. In nonrandomized studies conducted in the 1990s, LDCT scans were shown to detect more lung nodules and cancers than standard CXR in selected high-risk populations (e.g., age ≥ 60 years and a smoking history of ≥ 10 pack-years). Notably, up to 85% of the lung cancers discovered in these trials were classified as stage I disease and therefore considered potentially curable with surgical resection.

These data prompted the National Cancer Institute (NCI) to initiate the National Lung Screening Trial (NLST), a randomized study designed to determine if LDCT screening could reduce mortality from lung cancer in high-risk populations as compared with standard posterior anterior CXR. High-risk patients were defined as individuals between 55 and 74 years of age, with a ≥ 30 pack-year history of cigarette smoking; former smokers must have quit within the previous 15 years. Excluded from the trial were individuals with a previous lung cancer diagnosis, a history of hemoptysis, an unexplained weight loss of >15 lb in the preceding year, or a chest CT within 18 months of enrollment. A total of 53,454 persons were enrolled and randomized to annual screening yearly for 3 years (LDCT screening, n = 26,722; CXR screening, n = 26,732). Any noncalcified nodule measuring ≥ 4 mm in any diameter found on LDCT and CXR images with any noncalcified nodule or mass were classified as "positive." Participating radiologists had the option of not calling a final screen positive if a noncalcified nodule had been stable on the three screening examinations. Overall, 39.1% of participants in the LDCT group and 16% in the CXR group had at least one positive screening result. Of those who screened positive, the false-positive rate was 96.4% in the LDCT group and 94.5% in the CXR group. This was consistent across all three rounds. In the LDCT group, 1060 cancers were identified compared with 941 cancers in the CXR group (645 vs 572 per 100,000 person-years; relative risk [RR], 1.13; 95% confidence interval [CI], 1.03–1.23). Nearly twice as many stage IA cancers were detected in the LDCT group compared with the CXR group (40% vs 21%). The overall rates of lung cancer death were 247 and 309 deaths per 100,000 participants in the LDCT and CXR groups, respectively, representing a 20% reduction in lung cancer mortality in the LDCT-screened population (95% CI, 6.8–26.7%; p = .004). Compared with the CXR group, the rate of death in the LDCT group from *any* cause was reduced by 6.7% (95% CI, 1.2–13.6%; p = .02). The number needed to screen (NNTS) to prevent one lung cancer death was calculated to be 320.

The Nelson study was a second randomized trial comparing no screening to CT scans at baseline and in years 1, 3, and 5.5 in 13,195 men and 2594 women. Participants were 50–75 years of age and were current and former smokers with 10 years or less of cessation who smoked >15 cigarettes a day for >25 years or >10 cigarettes daily for >30 years. Participants were selected from four regions in the Netherlands or Belgium and were excluded if they were in moderate or bad self-reported health, were unable to climb two flights of stairs, had a body weight >140 kg, had a CT of the chest within the past year or a history of lung cancer <5 years ago or were still under treatment, or had current or past renal cell carcinoma, melanoma, or breast cancer. The hazard ratio for lung cancer mortality at 10 years was 0.74 (95% CI, 0.60–0.91; p = .003) and 0.61 (95% CI, 0.35–1.04; p = .0543) in men and women, respectively. These two trials have validated the use of annual CT scans for early detection of lung cancer in high-risk populations.

LDCT screening for lung cancer comes with known risks including a high rate of false-positive results, false-negative results, potential for unnecessary follow-up testing, radiation exposure, overdiagnosis, changes in anxiety level and quality of life, and substantial financial costs. By far, the biggest challenge confronting the use of CT screening is the high false-positive rate. False positives can have a substantial impact on patients through the expense and risk of unneeded further evaluation and emotional stress. The management of these patients usually consists of serial CT scans over time to see if the nodules grow, attempted fine-needle aspirates, or surgical resection. At \$300 per scan (NCI estimated cost), the outlay for initial LDCT alone could run into the billions of dollars annually, an expense that only further escalates when factoring in various downstream expenditures an individual might incur in the assessment of positive findings. A formal cost-effectiveness analysis of the NLST demonstrated differences between sex, age, and current smoking status and the method of follow-up. Despite some questions, LDCT screening has been recommended for all patients meeting criteria for enrollment on NLST. When discussing the option of LDCT screening, use of absolute risks rather than relative risks is helpful because studies indicate the public can process absolute terminology more effectively than relative risk projections. A useful

TABLE 78-2 The Benefits and Harms of LDCT Screening for Lung Cancer Based on NLST Data

	LDCT	CXR
Benefits: How did CT scans help compared to CXR?		
4 in 1000 fewer died from lung cancer	13 in 1000	17 in 1000
5 in 1000 fewer died from all causes	70 in 1000	75 in 1000
Harms: What problems did CT scans cause compared to CXR?		
223 in 1000 had at least 1 false alarm	365 in 1000	142 in 1000
18 in 1000 had a false alarm leading to an invasive procedure	25 in 1000	7 in 1000
2 in 1000 had a major complication from an invasive procedure	3 in 1000	1 in 1000

Abbreviations: CXR, chest x-ray; LDCT, low-dose computed tomography; NLST, National Lung Screening Trial.

Source: From S Woloshin: Cancer screening campaigns getting past uninformative persuasion. *N Engl J Med* 367:1167, 2012. Copyright © (2012) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

guide has been developed by the NCI to help patients and physicians assess the benefits and harms of LDCT screening for lung cancer (**Table 78-2**).

CLINICAL MANIFESTATIONS

Over half of all patients diagnosed with lung cancer present with locally advanced or metastatic disease at the time of diagnosis. The majority of patients present with signs, symptoms, or laboratory abnormalities that can be attributed to the primary lesion, local tumor growth, invasion or obstruction of adjacent structures, growth at distant metastatic sites, or a paraneoplastic syndrome (**Tables 78-3** and **78-4**). The prototypical lung cancer patient is a current or former smoker of either sex, usually in the seventh decade of life. A history of chronic cough with or without hemoptysis in a current or former smoker with chronic obstructive pulmonary disease (COPD) age 40 years or older should prompt a thorough investigation for lung cancer even in the face of a normal CXR. A persistent pneumonia without constitutional symptoms and unresponsive to repeated courses of antibiotics also should prompt an evaluation for the underlying cause. Lung cancer arising in a lifetime never smoker is more common in women and East Asians. Such patients also tend to be younger than their smoking counterparts at the time of diagnosis. The clinical presentation of lung cancer in never smokers tends to mirror that of current and former smokers.

Patients with central or endobronchial growth of the primary tumor may present with cough, hemoptysis, wheeze, stridor, dyspnea, or

TABLE 78-3 Presenting Signs and Symptoms of Lung Cancer

SYMPTOM AND SIGNS	RANGE OF FREQUENCY
Cough	8–75%
Weight loss	0–68%
Dyspnea	3–60%
Chest pain	20–49%
Hemoptysis	6–35%
Bone pain	6–25%
Clubbing	0–20%
Fever	0–20%
Weakness	0–10%
Superior vena cava obstruction	0–4%
Dysphagia	0–2%
Wheezing and stridor	0–2%

Source: Reproduced with permission from MA Beckles: Initial evaluation of the patient with lung cancer. Symptoms, sign, laboratory tests, and paraneoplastic syndromes. *Chest* 123:97, 2003.

TABLE 78-4 Clinical Findings Suggestive of Metastatic Disease

Symptoms elicited in history	<ul style="list-style-type: none"> Constitutional: weight loss >10 lb Musculoskeletal: pain Neurologic: headaches, syncope, seizures, extremity weakness, recent change in mental status
Signs found on physical examination	<ul style="list-style-type: none"> Lymphadenopathy (>1 cm) Hoarseness, superior vena cava syndrome Bone tenderness Hepatomegaly (>13 cm span) Focal neurologic signs, papilledema Soft-tissue mass
Routine laboratory tests	<ul style="list-style-type: none"> Hematocrit, <40% in men; <35% in women Elevated alkaline phosphatase, GGT, SGOT, and calcium levels

Abbreviations: GGT, gamma-glutamyltransferase; SGOT, serum glutamic-oxaloacetic transaminase.

Source: Reproduced with permission from GA Silvestri et al: The noninvasive staging of non-small cell lung cancer. *Chest* 123:147S, 2003.

postobstructive pneumonia. Peripheral growth of the primary tumor may cause pain from pleural or chest wall involvement, dyspnea on a restrictive basis, and symptoms of a lung abscess resulting from tumor cavitation. Regional spread of tumor in the thorax (by contiguous growth or by metastasis to regional lymph nodes) may cause tracheal obstruction, esophageal compression with dysphagia, recurrent laryngeal nerve paralysis with hoarseness, phrenic nerve palsy with elevation of the hemidiaphragm and dyspnea, and sympathetic nerve paralysis with Horner's syndrome (enophthalmos, ptosis, miosis, and anhidrosis). Malignant pleural effusions can cause pain, dyspnea, or cough. Pancoast (or superior sulcus tumor) syndromes result from local extension of a tumor growing in the apex of the lung with involvement of the eighth cervical and first and second thoracic nerves, and present with shoulder pain that characteristically radiates in the ulnar distribution of the arm, often with radiologic destruction of the first and second ribs. Often Horner's syndrome and Pancoast syndrome coexist. Other problems of regional spread include superior vena cava syndrome from vascular obstruction; pericardial and cardiac extension with resultant tamponade, arrhythmia, or cardiac failure; lymphatic obstruction with resultant pleural effusion; and lymphangitic spread through the lungs with hypoxemia and dyspnea. In addition, lung cancer can spread transbronchially, producing tumor growth along multiple alveolar surfaces with impairment of gas exchange, respiratory insufficiency, dyspnea, hypoxemia, and sputum production. Constitutional symptoms may include anorexia, weight loss, weakness, fever, and night sweats. Apart from the brevity of symptom duration, these parameters fail to clearly distinguish SCLC from NSCLC or even from neoplasms metastatic to lungs.

Extrathoracic metastatic disease is found at autopsy in >50% of patients with squamous carcinoma, 80% of patients with adenocarcinoma and large-cell carcinoma, and >95% of patients with SCLC. Approximately one-third of patients present with symptoms as a result of distant metastases. Lung cancer metastases may occur in virtually every organ system, and the site of metastatic involvement largely determines other symptoms. Patients with brain metastases may present with headache, nausea and vomiting, seizures, or neurologic deficits. Patients with bone metastases may present with pain, pathologic fractures, or spinal cord compression. The latter may also occur with epidural metastases. Individuals with bone marrow invasion may present with cytopenias or leukoerythroblastosis. Those with liver metastases may present with hepatomegaly, right upper quadrant pain, fever, anorexia, and weight loss. Liver dysfunction and biliary obstruction are rare. Adrenal metastases are common but rarely cause pain or adrenal insufficiency unless they are large.

Paraneoplastic syndromes are common in patients with lung cancer, especially those with SCLC, and may be the presenting finding or the first sign of recurrence. In addition, paraneoplastic syndromes may mimic metastatic disease and, unless detected, lead to inappropriate

palliative rather than curative treatment. Often the paraneoplastic syndrome may be relieved with successful treatment of the tumor. In some cases, the pathophysiology of the paraneoplastic syndrome is known, particularly when a hormone with biologic activity is secreted by a tumor. However, in many cases, the pathophysiology is unknown. Systemic symptoms of anorexia, cachexia, weight loss (seen in 30% of patients), fever, and suppressed immunity are paraneoplastic syndromes of unknown etiology or at least not well defined. Weight loss >10% of total body weight is considered a bad prognostic sign. Endocrine syndromes are seen in 12% of patients; hypercalcemia resulting from ectopic production of parathyroid hormone (PTH) or, more commonly, PTH-related peptide is the most common life-threatening metabolic complication of malignancy, primarily occurring with squamous cell carcinomas of the lung. Clinical symptoms include nausea, vomiting, abdominal pain, constipation, polyuria, thirst, and altered mental status.

Hyponatremia may be caused by the syndrome of inappropriate secretion of antidiuretic hormone (SIADH) or possibly atrial natriuretic peptide (ANP) (Chap. 93). SIADH resolves within 1–4 weeks of initiating chemotherapy in the vast majority of cases. During this period, serum sodium can usually be managed and maintained above 128 mEq/L via fluid restriction. Demeocycline can be a useful adjunctive measure when fluid restriction alone is insufficient. Vasopressin receptor antagonists like tolvaptan also have been used in the management of SIADH. However, the use of tolvaptan has significant limitations including liver injury and overly rapid correction of the hyponatremia, which can lead to irreversible neurologic injury. Likewise, the cost of tolvaptan may be prohibitive (as high as \$300 per tablet in some areas). Of note, patients with ectopic ANP may have worsening hyponatremia if sodium intake is not concomitantly increased. Accordingly, if hyponatremia fails to improve or worsens after 3–4 days of adequate fluid restriction, plasma levels of ANP should be measured to determine the causative syndrome.

Ectopic secretion of ACTH by SCLC and pulmonary carcinoids usually results in additional electrolyte disturbances, especially hypokalemia, rather than the changes in body habitus that occur in Cushing's syndrome from a pituitary adenoma (Chap. 93). Treatment with standard medications, such as metyrapone and ketoconazole, is largely ineffective due to extremely high cortisol levels. The most effective strategy for management of the Cushing's syndrome is effective treatment of the underlying SCLC. Bilateral adrenalectomy may be considered in extreme cases.

Skeletal-connective tissue syndromes include clubbing in 30% of cases (usually NSCLCs) and hypertrophic primary osteoarthropathy in 1–10% of cases (usually adenocarcinomas). Patients may develop periostitis, causing pain, tenderness, and swelling over the affected bones and a positive bone scan. Neurologic-myopathic syndromes are seen in only 1% of patients but are dramatic and include the myasthenic Eaton-Lambert syndrome and retinal blindness with SCLC, whereas peripheral neuropathies, subacute cerebellar degeneration, cortical degeneration, and polymyositis are seen with all lung cancer types. Many of these are caused by autoimmune responses such as the development of anti-voltage-gated calcium channel antibodies in Eaton-Lambert syndrome. Patients with this disorder present with proximal muscle weakness, usually in the lower extremities, occasional autonomic dysfunction, and rarely, cranial nerve symptoms or involvement of the bulbar or respiratory muscles. Depressed deep tendon reflexes are frequently present. In contrast to patients with myasthenia gravis, strength improves with serial effort. Some patients who respond to chemotherapy will have resolution of the neurologic abnormalities. Thus, chemotherapy is the initial treatment of choice. Paraneoplastic encephalomyelitis and sensory neuropathies, cerebellar degeneration, limbic encephalitis, and brainstem encephalitis occur in SCLC in association with a variety of antineuronal antibodies such as anti-Hu, anti-CRMP5, and ANNA-3. Paraneoplastic cerebellar degeneration may be associated with anti-Hu, anti-Yo, or P/Q calcium channel autoantibodies. Coagulation or thrombotic or other hematologic manifestations occur in 1–8% of patients and include migratory venous thrombophlebitis (Trousseau's syndrome), nonbacterial thrombotic (marantic) endocarditis with arterial emboli, and disseminated

intravascular coagulation with hemorrhage, anemia, granulocytosis, and leukoerythroblastosis. Thrombotic disease complicating cancer is usually a poor prognostic sign. Cutaneous manifestations such as dermatomyositis and acanthosis nigricans are uncommon (1%), as are the renal manifestations of nephrotic syndrome and glomerulonephritis ($\leq 1\%$).

DIAGNOSING LUNG CANCER

Tissue sampling is required to confirm a diagnosis in all patients with suspected lung cancer. In patients with suspected metastatic disease, a biopsy of a distant site of disease is preferred for tissue confirmation. Given the greater emphasis placed on molecular and PD-L1 testing for NSCLC patients, a core biopsy is preferred to ensure adequate tissue for analysis. Tumor tissue may be obtained via minimally invasive techniques such as bronchial or transbronchial biopsy during fiberoptic bronchoscopy, by fine-needle aspiration (FNA) or percutaneous biopsy using image guidance, or via endobronchial ultrasound (EBUS)-guided biopsy. Depending on the location, lymph node sampling may occur via transesophageal endoscopic ultrasound (EUS)-guided biopsy, EBUS-guided biopsy, or blind biopsy. In patients with suspected metastatic disease, a diagnosis may be confirmed by bronchoscopy, percutaneous biopsy of a soft tissue mass, lytic bone lesion, bone marrow, pleural or liver lesion, or an adequate cell block obtained from a malignant pleural effusion. In patients with a suspected malignant pleural effusion, if the initial thoracentesis is negative, a repeat thoracentesis is warranted. Although the majority of pleural effusions are due to malignant disease, particularly if they are exudative or bloody, some may be parapneumonic. In the absence of distant disease, such patients should be considered for possible curative treatment.

The diagnostic yield of any biopsy depends on several factors including location (accessibility) of the tumor, tumor size, tumor type, and technical aspects of the diagnostic procedure including the experience level of the bronchoscopist and pathologist. In general, central lesions such as squamous cell carcinomas, small-cell carcinomas, or endobronchial lesions such as carcinoid tumors are more readily diagnosed by bronchoscopic examination, whereas peripheral lesions such as adenocarcinomas and large-cell carcinomas are more amenable to transthoracic biopsy.

Sputum cytology is inexpensive and noninvasive but has a lower yield than other specimen types due to poor preservation of the cells and more variability in acquiring a good-quality specimen. The yield for sputum cytology is highest for larger and centrally located tumors such as squamous cell carcinoma and small-cell carcinoma histology. The specificity for sputum cytology averages close to 100%, although sensitivity is generally <70%. The accuracy of sputum cytology improves with increased numbers of specimens analyzed. Consequently, analysis of at least three sputum specimens is recommended. However, the quality of the specimen may not be adequate for histologic subclassification and PD-L1 and molecular testing.

STAGING LUNG CANCER

Lung cancer staging consists of two parts: first, a determination of the location of the tumor and possible metastatic sites (anatomic staging), and second, an assessment of a patient's ability to withstand various antitumor treatments (physiologic staging). All patients with lung cancer should have a complete history and physical examination, with evaluation of all other medical problems, determination of performance status, and history of weight loss. Staging with regard to a patient's potential for surgical resection is principally applicable to NSCLC.

■ ANATOMIC STAGING OF PATIENTS WITH LUNG CANCER

The accurate staging of patients with NSCLC is essential for determining the appropriate treatment in patients with resectable disease and for avoiding unnecessary surgical procedures in patients with advanced disease. All patients with NSCLC should undergo initial radiographic imaging with CT scan, positron emission tomography (PET), or preferably CT-PET. PET scanning attempts to identify sites of malignancy based on glucose metabolism by measuring the uptake

of ^{18}F -fluorodeoxyglucose (FDG). Rapidly dividing cells, presumably in the lung tumors, will preferentially take up ^{18}F -FDG and appear as a “hot spot.” To date, PET has been mostly used for staging and detection of metastases in lung cancer and in the detection of nodules >15 mm in diameter. Combined ^{18}F -FDG PET-CT imaging has been shown to improve the accuracy of staging in NSCLC compared to visual correlation of PET and CT or either study alone. CT-PET has been found to be superior in identifying pathologically enlarged mediastinal lymph nodes and extrathoracic metastases. A standardized uptake value (SUV) of >2.5 on PET is highly suspicious for malignancy. False negatives can be seen in diabetes, in lesions <8 mm, and in slow-growing tumors (e.g., carcinoid tumors or well-differentiated adenocarcinoma). False positives can be seen in certain infections and granulomatous disease (e.g., tuberculosis). Thus, PET should never be used alone to diagnose lung cancer, mediastinal involvement, or metastases. Confirmation with tissue biopsy is required. For brain metastases, magnetic resonance imaging (MRI) is the most effective method. MRI can also be useful in selected circumstances, such as superior sulcus tumors to rule out brachial plexus involvement, but in general, MRI does not play a major role in NSCLC staging.

In patients with NSCLC, the following are contraindications to potential curative resection: extrathoracic metastases, superior vena cava syndrome, vocal cord and, in most cases, phrenic nerve paralysis, malignant pleural effusion, cardiac tamponade, tumor within 2 cm of the carina (potentially curable with combined chemoradiotherapy), metastasis to the contralateral lung, metastases to supraclavicular lymph nodes, contralateral mediastinal node metastases (potentially curable with combined chemoradiotherapy), and involvement of the main pulmonary artery. In situations where it will make a difference in treatment, abnormal scan findings require tissue confirmation of malignancy so that patients are not precluded from having potentially curative therapy.

The best predictor of metastatic disease remains a careful history and physical examination. If signs, symptoms, or findings from the physical examination suggest the presence of malignancy, then sequential imaging starting with the most appropriate study should be performed. If the findings from the clinical evaluation are negative, then imaging studies beyond CT-PET are unnecessary and the search for metastatic disease is complete. In patients in whom distant metastatic disease has been ruled out, lymph node status needs to be assessed via minimally invasive techniques such as those mentioned above and/or invasive techniques such as mediastinoscopy, mediastinotomy, thoracoscopy, or thoracotomy. Approximately one-quarter to one-half of patients diagnosed with NSCLC will have mediastinal lymph node metastases at the time of diagnosis. Lymph node sampling is recommended in all patients with enlarged nodes detected by CT or PET scan and in patients with large tumors or tumors occupying the inner third of the lung. The extent of mediastinal lymph node involvement is

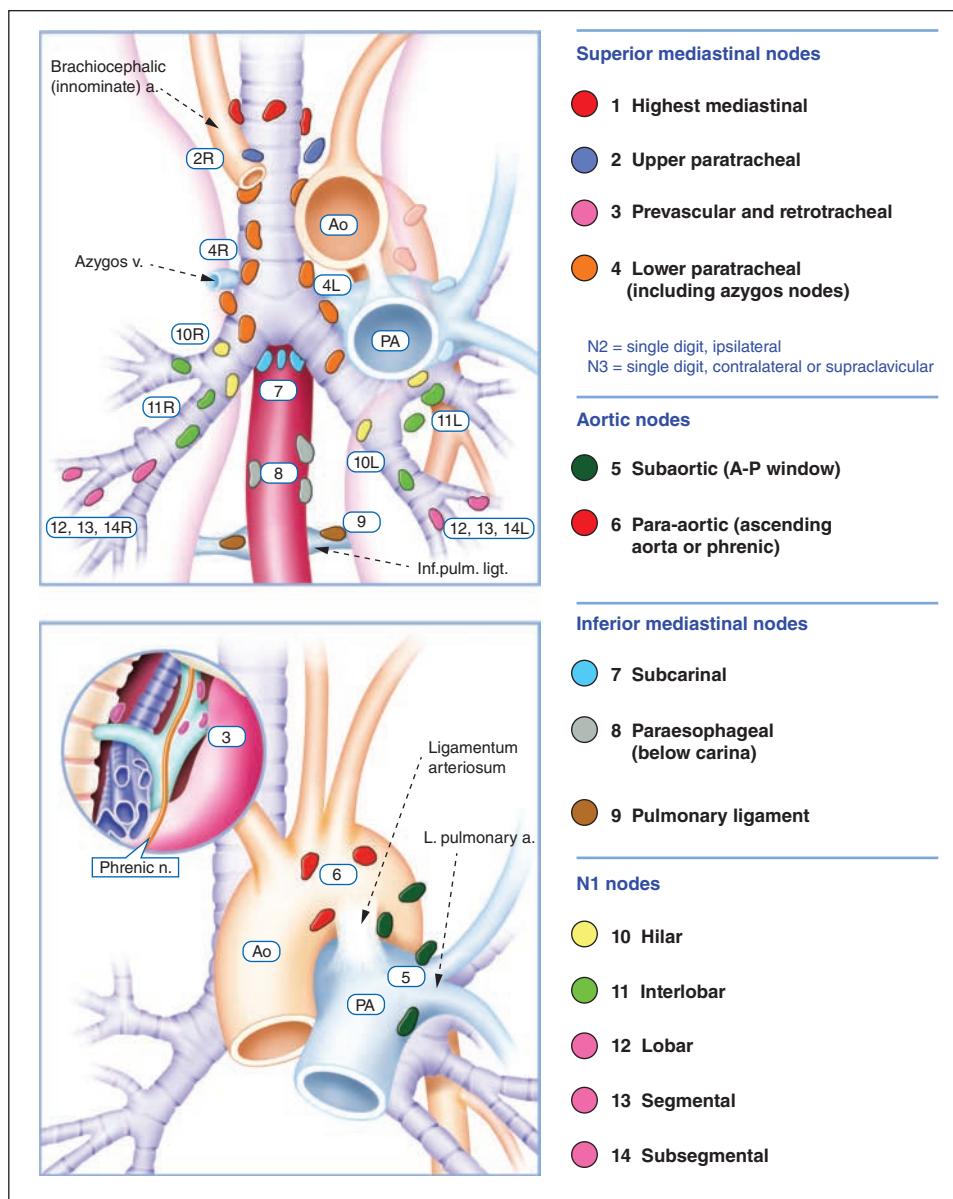


FIGURE 78-3 Lymph node stations in staging non-small-cell lung cancer. The International Association for the Study of Lung Cancer (IASLC) lymph node map, including the proposed grouping of lymph node stations into “zones” for the purposes of prognostic analyses. a., artery; Ao, aorta; Inf. pulm. ligt., inferior pulmonary ligament; n., nerve; PA, pulmonary artery; v., vein.

important in determining the appropriate treatment strategy: surgical resection followed by adjuvant chemotherapy versus combined chemoradiation followed by immunotherapy (durvalumab) (see below). A standard nomenclature for referring to the location of lymph nodes involved with lung cancer has evolved (Fig. 78-3).

In SCLC patients, current staging recommendations include a PET-CT scan and MRI of the brain (positive in 10% of asymptomatic patients). Bone marrow biopsies and aspirations are rarely performed now given the low incidence of isolated bone marrow metastases. Confirmation of metastatic disease, ipsilateral or contralateral lung nodules, or metastases beyond the mediastinum may be achieved by the same modalities recommended earlier for patients with NSCLC.

If a patient has signs or symptoms of spinal cord compression (pain, weakness, paralysis, urinary retention), a spinal CT or MRI scan should be performed. If metastases are evident on imaging, a neurosurgeon should be consulted for possible palliative surgical resection and/or a radiation oncologist should be consulted for palliative radiotherapy to the site of compression. If signs or symptoms of leptomeningeal disease develop at any time in a patient with lung cancer, an MRI of the brain and spinal cord should be performed, as well as a spinal tap, for

detection of malignant cells. If the spinal tap is negative, a repeat spinal tap should be considered. There is currently no approved therapy for the treatment of leptomeningeal disease.

■ STAGING SYSTEM FOR NON-SMALL-CELL LUNG CANCER

The tumor-node-metastasis (TNM) international staging system provides useful prognostic information and is used to stage all patients with NSCLC. The various T (tumor size), N (regional node involvement), and M (presence or absence of distant metastasis) stages are combined to form different stage groups (Tables 74-5 and 74-6). The eighth edition of the TNM staging system went into effect in 2018. T1 tumors are divided into tumors ≤ 1 cm (T1a), >1 cm and ≤ 2 cm (T1b), and >2 cm and ≤ 3 cm (T1c). T2 tumors are those that are >3 cm but ≤ 5 cm, involve the visceral pleura or main bronchus, or are associated with atelectasis; T2a tumors are >3 cm and ≤ 4 cm, and T2b are >4 cm and ≤ 5 cm. T3 tumors are >5 cm and ≤ 7 cm. T3 tumors also include tumors with invasion into local structures such as the chest wall and diaphragm and with additional nodules in the same lobe. T4 tumors include tumors >7 cm or tumors of any size with invasion into mediastinum, heart, great vessels, trachea, esophagus, or spine

TABLE 78-5 TNM Staging System for Lung Cancer (Eighth Edition)

Primary Tumor (T)	
T1	Tumor ≤ 3 cm diameter, surrounded by lung or visceral pleura, without invasion more proximal than lobar bronchus
T1mi	Minimally invasive adenocarcinoma (pure lepidic pattern, <3 cm in greatest dimension and <5 mm invasion)—T1a (size <1 cm)—T1b (1 cm $<$ size <2 cm)—T1c (2 cm $<$ size <3 cm)
T2	Tumor >3 cm but ≤ 7 cm, or tumor with any of the following features: Involves main bronchus ≥ 2 cm distal to carina Invades visceral pleura Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung
T2a	Tumor >3 cm but ≤ 5 cm
T2b	Tumor >5 cm but ≤ 7 cm
T3	Tumor >7 cm or any of the following: Directly invades any of the following: chest wall, diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium, main bronchus <2 cm from carina (without involvement of carina) Atelectasis or obstructive pneumonitis of the entire lung Separate tumor nodules in the same lobe
T4	Tumor of any size that invades the mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, or carina, or with separate tumor nodules in a different ipsilateral lobe
Nodal Stage (N)	
N0	No regional lymph node metastases
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)
Metastases (M)	
M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumor nodule(s) in a contralateral lobe; tumor with pleural nodules or malignant pleural or pericardial effusion
M1b	Distant metastasis (in extrathoracic organs)
M1c	

Abbreviation: M1b, distant metastasis in single organ outside chest; M1c, multiple extrathoracic metastases to one or more organs; TNM, tumor-node-metastasis.

TABLE 78-6 TNM Stage Groupings, Eighth Edition

Stage IA1	T1a	N0	M0
Stage IA2	T1b	N0	M0
Stage IA3	T1c	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0	M0
Stage IIB	T1a-T2b T3	N1 N0	M0 M0
Stage IIIA	T1-2b T3	N2 N1	M0 M0
	T4	N0/N1	M0
Stage IIIB	T1-2b T3/T4	N3 N0/N1 N3	M0 M0 M0
Stage IVA	Any T	Any N	M1a/M1b
Stage IVB	Any T	Any N	M1c

or with multiple nodules in the ipsilateral lung. Lymph node staging depends on metastasis to ipsilateral pulmonary or hilar nodes (N1), mediastinal or subcarinal nodes (N2), or contralateral mediastinal, hilar, or supraclavicular nodes (N3). Patients with metastasis may be classified as M1a (malignant pleural or pericardial effusion, pleural nodules, or nodules in the contralateral lung), M1b (single distant metastasis to a single organ; e.g., bone, liver, adrenal, or brain metastasis), or M1c (multiple metastases to a single organ or metastases to multiple organs). The effect of stage on survival is illustrated in Fig. 78-4. Approximately 15% of patients have localized disease that can be treated with curative attempt (surgery or radiotherapy), about a quarter have local or regional disease that may or may not be amenable to a curative attempt, and half have metastatic disease at the time of diagnosis. In 10%, the extent of disease is undefined.

■ STAGING SYSTEM FOR SMALL-CELL LUNG CANCER

In patients with SCLC, it is now recommended that both the Veterans Administration system and the American Joint Committee on Cancer/International Union Against Cancer eighth edition system (TNM) be used to classify the tumor stage. The Veterans Administration system is a distinct two-stage system dividing patients into those with limited- or extensive-stage disease. Patients with limited-stage disease (LD) have cancer that is confined to the ipsilateral hemithorax and can be encompassed within a tolerable radiation port. Thus, contralateral supraclavicular nodes, recurrent laryngeal nerve involvement, and superior vena caval obstruction can all be part of LD. Patients with extensive-stage disease (ED) have overt metastatic disease by imaging or physical examination. Cardiac tamponade, malignant pleural effusion, and bilateral pulmonary parenchymal involvement generally qualify disease as ED, because the involved organs cannot be encompassed safely or effectively within a single radiation therapy port. Sixty to 70% of patients are diagnosed with ED at presentation. The TNM staging system is preferred in the rare SCLC patient presenting with what appears to be clinical stage I disease (see above).

■ PHYSIOLOGIC STAGING

Patients with lung cancer often have other comorbid conditions related to smoking including cardiovascular disease and COPD. To improve their preoperative condition, correctable problems (e.g., anemia, electrolyte and fluid disorders, infections, cardiac disease, and arrhythmias) should be addressed, appropriate chest physical therapy should be instituted, and patients should be encouraged to stop smoking. Patients with a forced expiratory volume in 1 s (FEV₁) of >2 L or $>80\%$ of predicted can tolerate a pneumonectomy, and those with an FEV₁ >1.5 L have adequate reserve for a lobectomy. In patients with borderline lung function but a resectable tumor, cardiopulmonary exercise testing could be performed as part of the physiologic evaluation. This test allows an estimate of the maximal oxygen consumption ($V_{O_{2max}}$). A $V_{O_{2max}}$

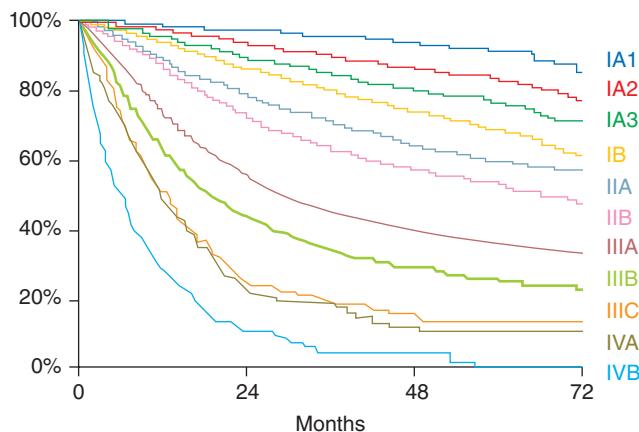


FIGURE 78-4 Influence of non-small-cell lung cancer stage on survival.

Stage	24 months	60 months
IA1	97%	92%
IA2	94%	83%
IA3	90%	77%
IB	87%	68%
IIA	79%	60%
IIB	72%	53%
IIIA	55%	36%
IIIB	44%	26%
IIIC	24%	13%
IVA	23%	10%
IVB	10%	0%

<15 mL/kg·min) predicts for a higher risk of postoperative complications. Patients deemed unable to tolerate lobectomy or pneumonectomy from a pulmonary functional standpoint may be candidates for more limited resections, such as wedge or anatomic segmental resection, although such procedures are associated with significantly higher rates of local recurrence and a trend toward decreased overall survival. All patients should be assessed for cardiovascular risk using American College of Cardiology and American Heart Association guidelines. A myocardial infarction within the past 3 months is a contraindication to thoracic surgery because 20% of patients will die of reinfarction. An infarction in the past 6 months is a relative contraindication. Other major contraindications include uncontrolled arrhythmias, an FEV₁ of <1 L, CO₂ retention (resting PCO₂ >45 mmHg), DLCO <40%, and severe pulmonary hypertension.

TREATMENT

Non-Small-Cell Lung Cancer

The overall treatment approach to patients with NSCLC is shown in Fig. 78-5.

OCCULT AND STAGE 0 CARCINOMAS

Patients with severe atypia on sputum cytology have an increased risk of developing lung cancer compared to those without atypia. In the uncommon circumstance where malignant cells are identified in a sputum or bronchial washing specimen but the chest imaging appears normal (TX tumor stage), the lesion must be localized. More than 90% of tumors can be localized by meticulous examination of the bronchial tree with a fiberoptic bronchoscope under general anesthesia and collection of a series of differential brushings and biopsies. Surgical resection following bronchoscopic localization has been shown to improve survival compared to no treatment. Close follow-up of these patients is indicated because of the high incidence of second primary lung cancers (5% per patient per year).

SOLITARY PULMONARY NODULE AND "GROUND-GLASS" OPACITIES

A solitary pulmonary nodule is defined as an x-ray density completely surrounded by normal aerated lung with circumscribed margins, of any shape, usually 1–6 cm in greatest diameter. The approach to a patient with a solitary pulmonary nodule is based on an estimate of the probability of cancer, determined according to the patient's smoking history, age, and characteristics on imaging (Table 78-7). Prior CXRs and CT scans should be obtained if available for comparison. A PET scan may be useful if the lesion is greater than 7–8 mm in diameter. If no diagnosis is apparent, Mayo investigators reported that clinical characteristics (age, cigarette

smoking status, and prior cancer diagnosis) and three radiologic characteristics (nodule diameter, spiculation, and upper lobe location) were independent predictors of malignancy. At present, only two radiographic criteria are thought to predict the benign nature of a solitary pulmonary nodule: lack of growth over a period >2 years and certain characteristic patterns of calcification. Calcification alone, however, does not exclude malignancy; a dense central nidus, multiple punctate foci, and "bull's eye" (granuloma) and "popcorn ball" (hamartoma) calcifications are highly suggestive of a benign lesion. In contrast, a relatively large lesion, lack of or asymmetric calcification, chest symptoms, associated atelectasis, pneumonitis, or growth of the lesion revealed by comparison with an old x-ray or CT scan or a positive PET scan may be suggestive of a malignant process and warrant further attempts to establish a histologic diagnosis. An algorithm for assessing these lesions is shown in Fig. 78-6.

Since the advent of screening CTs, small "ground-glass" opacities (GGOs) have often been observed, particularly as the increased sensitivity of CTs enables detection of smaller lesions. Many of these GGOs, when biopsied, are found to be atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), or minimally invasive adenocarcinoma (MIA). AAH is usually a nodule of <5 mm and is minimally hazy, also called nonsolid or ground glass (i.e., hazy slightly increased attenuation, no solid component, and preservation of bronchial and vascular margins). On thin-section CT, AIS is usually a nonsolid nodule and tends to be slightly more opaque than AAH. MIA is mainly solid, usually with a small (<5 mm) central solid component. However, overlap exists among the imaging features of the preinvasive and minimally invasive lesions in the lung adenocarcinoma spectrum. Lepidic adenocarcinomas are usually solid but may be nonsolid. Likewise, the small invasive adenocarcinomas also are usually solid but may exhibit a small nonsolid component.

MANAGEMENT OF STAGES I AND II NSCLC

Surgical Resection of Stage I and II NSCLC Surgical resection, ideally by an experienced thoracic surgeon, is the treatment of choice for patients with clinical stage I and II NSCLC who are able to tolerate the procedure. Operative mortality rates for patients resected by thoracic or cardiothoracic surgeons are lower compared to general surgeons. Moreover, survival rates are higher in patients who undergo resection in facilities with a high surgical volume compared to those performing fewer than 70 procedures per year, even though the higher-volume facilities often serve older and less socioeconomically advantaged populations. The improvement in survival is most evident in the immediate postoperative period. In patients with stage I NSCLC, lobectomy is superior to wedge

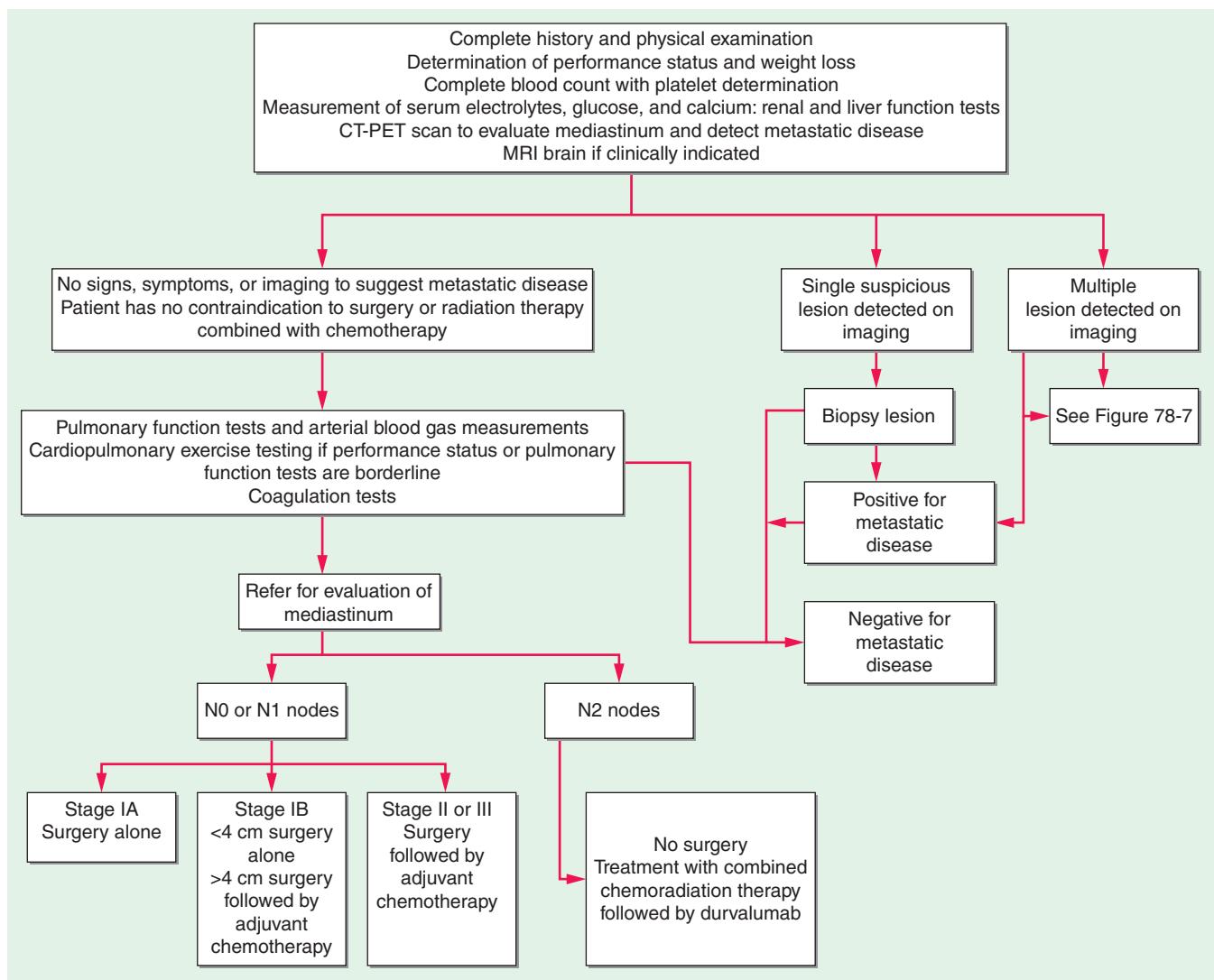


FIGURE 78-5 Algorithm for management of non-small-cell lung cancer. MRI, magnetic resonance imaging; PET, positron emission tomography.

resection with respect to rates of local recurrence. There is also a trend toward improvement in overall survival. In patients with comorbidities, compromised pulmonary reserve, and small peripheral lesions, a limited resection, wedge resection, or segmentectomy (potentially by video-assisted thoracoscopic surgery) may be reasonable surgical options. Pneumonectomy is reserved for patients with central tumors and should be performed only in patients with excellent pulmonary reserve. The 5-year survival rates are 68–92% for patients with stage I NSCLC and 53–60% for patients with stage II NSCLC.

TABLE 78-7 Assessment of Risk of Cancer in Patients with Solitary Pulmonary Nodules

VARIABLE	RISK		
	LOW	INTERMEDIATE	HIGH
Diameter (cm)	<1.5	1.5–2.2	≥2.3
Age (years)	<45	45–60	>60
Smoking status	Never smoker	Current smoker (<20 cigarettes/d)	Current smoker (>20 cigarettes/d)
Smoking cessation status	Quit ≥7 years ago or quit	Quit <7 years ago	Never quit
Characteristics of nodule margins	Smooth	Scalloped	Corona radiata or spiculated

Source: From D Ost et al: The solitary pulmonary nodule. N Engl J Med 348:2535, 2003. Copyright © (2003) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

Accurate pathologic staging requires adequate segmental, hilar, and mediastinal lymph node sampling. Ideally, this includes a mediastinal lymph node dissection. On the right side, mediastinal stations 2R, 4R, 7, 8R, and 9R should be dissected; on the left side, stations 5, 6, 7, 8L, and 9L should be dissected. Hilar lymph nodes are typically resected and sent for pathologic review, although it is helpful to specifically dissect and label level 10 lymph nodes when possible. On the left side, level 2 and sometimes level 4 lymph nodes are generally obscured by the aorta. Although the therapeutic benefit of nodal dissection versus nodal sampling is controversial, a pooled analysis of three trials involving patients with stages I to IIIA NSCLC demonstrated a superior 4-year survival in patients undergoing resection and a complete mediastinal lymph node dissection compared with lymph node sampling. Moreover, complete mediastinal lymphadenectomy added little morbidity to a pulmonary resection for lung cancer when carried out by an experienced thoracic surgeon.

Radiation Therapy in Stages I and II NSCLC There is currently no role for postoperative radiation therapy in patients following resection of stage I or II NSCLC with negative margins. However, patients with stage I and II disease who either refuse or are not suitable candidates for surgery should be considered for radiation therapy with *curative* intent. Stereotactic body radiation therapy (SBRT) is a technique used to treat patients with isolated pulmonary nodules (≤5 cm) who are not candidates for or refuse surgical resection. Treatment is typically administered in three to five

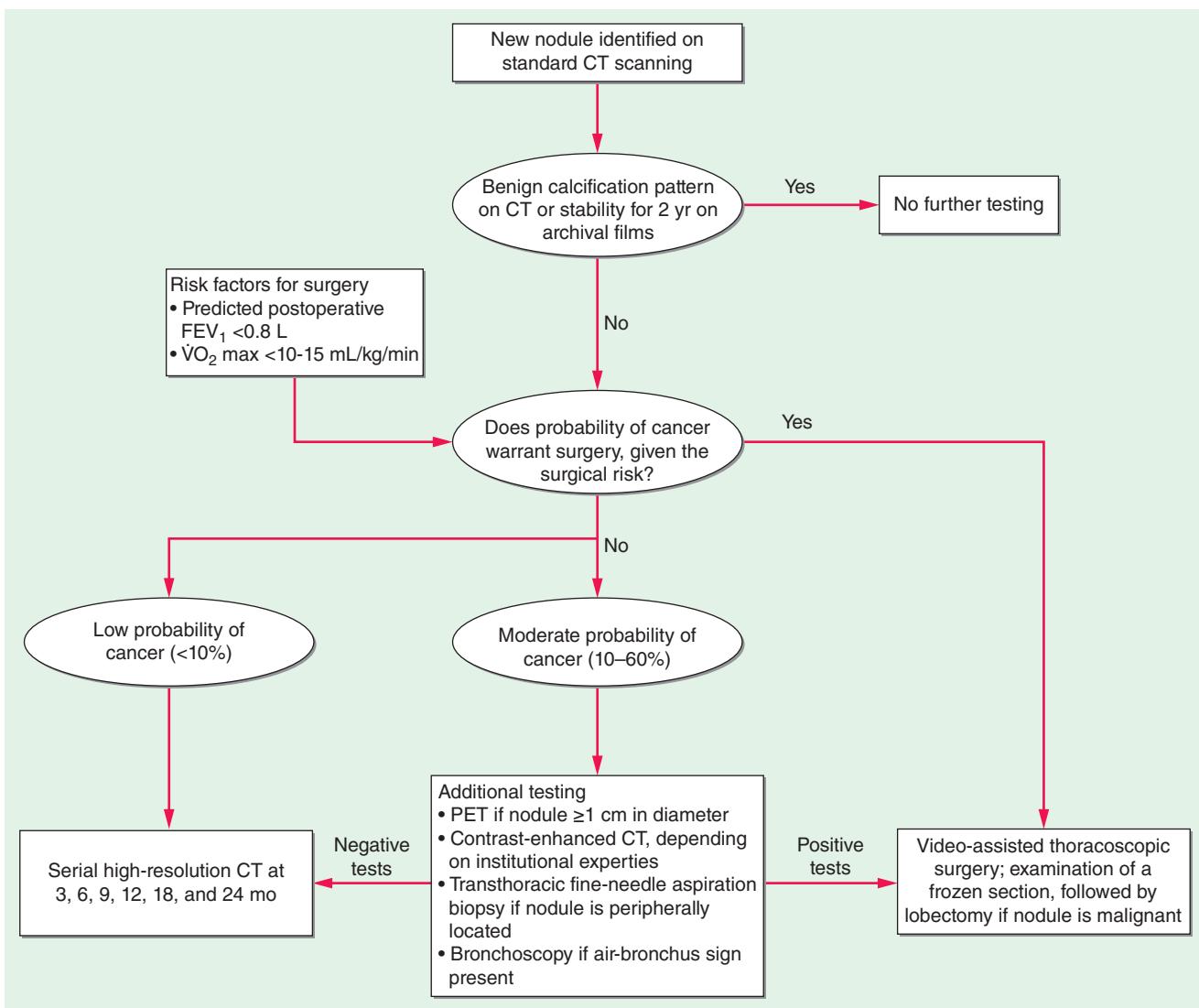


FIGURE 78-6 Approach to the solitary pulmonary nodule. FEV₁, forced expiratory volume in 1 s; PET, positron emission tomography.

fractions delivered over 1–2 weeks. In uncontrolled studies, disease control rates are >90%, and 5-year survival rates of up to 60% have been reported with SBRT. By comparison, survival rates typically range from 13 to 39% in patients with stage I or II NSCLC treated with standard external-beam radiotherapy. Cryoablation is another technique occasionally used to treat small, isolated tumors (i.e., ≤3 cm). However, very little data exist on long-term outcomes with this technique.

Chemotherapy in Stages I and II NSCLC Although a landmark meta-analysis of cisplatin-based adjuvant chemotherapy trials in patients with resected stages I to IIIA NSCLC (the Lung Adjuvant Cisplatin Evaluation [LACE] Study) demonstrated a 5.4% improvement in 5-year survival for adjuvant chemotherapy compared to surgery alone, the survival benefit was seemingly confined to patients with stage II or III disease (Table 78-8). By contrast, survival was actually worsened in stage IA patients with the application of adjuvant therapy. In stage IB, there was a modest improvement in survival of questionable clinical significance. Adjuvant chemotherapy was also detrimental in patients with poor performance status (Eastern Cooperative Oncology Group [ECOG] performance status = 2). These data suggest that adjuvant chemotherapy is best applied in patients with resected stage II or III NSCLC. There is no apparent role for adjuvant chemotherapy in patients with resected stage IA or IB NSCLC. A possible exception to the prohibition of adjuvant therapy in this setting is the stage IB patient with a resected

lesion ≥4 cm. Osimertinib, an EGFR TKI, demonstrated improved disease-free survival for patients with EGFR mutation (exon 19 deletion or L858R)-positive NSCLC treated for 3 years following chemotherapy. However, the effect on overall survival is unknown.

As with any treatment recommendation, the risks and benefits of adjuvant chemotherapy should be considered on an individual patient basis. If a decision is made to proceed with adjuvant chemotherapy, in general, treatment should be initiated 6–12 weeks after surgery, assuming the patient has fully recovered, and should be administered for no more than four cycles. Although cisplatin-based chemotherapy is the preferred treatment regimen, carboplatin can be substituted for cisplatin in patients who are unlikely to tolerate cisplatin for reasons such as reduced renal function, presence of neuropathy, or hearing impairment. A large cooperative group trial compared cisplatin-based chemotherapy with vinorelbine, pemetrexed, gemcitabine, or docetaxel with or without antiangiogenic therapy. While adding antiangiogenic therapy to platinum-based chemotherapy offered no benefit, the trial also demonstrated no difference in survival among the four chemotherapy regimens. Therefore, no specific chemotherapy regimen is considered superior in this setting, and treatment selection may be based on cost and patient comorbidities.

Neoadjuvant chemotherapy, which is the application of chemotherapy administered before an attempted surgical resection, has been advocated by some experts on the assumption that such an approach will more effectively extinguish occult micrometastases compared to

TABLE 78-8 Adjuvant Chemotherapy Trials in Non-Small-Cell Lung Cancer

TRIAL	STAGE	TREATMENT	NO. OF PATIENTS	5-YEAR SURVIVAL (%)	P
IALT	I–III	Cisplatin-based	932	44.5	<.03
		Control	835	40.4	
BR10	IB–II	Cisplatin + vinorelbine	242	69	.03
		Control	240	54	
ANITA	IB–IIIA	Cisplatin + vinorelbine	407	60	.017
		Control	433	58	
ALPI	I–III	MVP	548	50	.49
		Control	540	45	
BLT	I–III	Cisplatin-based	192	60	.90
		Control	189	58	
CALGB	IB	Carboplatin + paclitaxel	173	59	.10
			171	57	
ECOG1505	IB > 4c – IIIA	Cisplatin – based	749	NR	.90
		+ bevacizumab	752	NR	

Abbreviations: ALPI, Adjuvant Lung Cancer Project Italy; ANITA, Adjuvant Navelbine International Trialist Association; BLT, Big Lung Trial; CALGB, Cancer and Lung Cancer Group B; ECOG, Eastern Cooperative Oncology Group; IALT, International Adjuvant Lung Cancer Trial; MVP, mitomycin, vindesine, and cisplatin; NR, not reported.

postoperative chemotherapy. In addition, it is thought that preoperative chemotherapy might render an inoperable lesion resectable. A meta-analysis of 15 randomized controlled trials involving more than 2300 patients with stage I to III NSCLC suggested there may be a modest 5-year survival benefit (i.e., ~5%) that is virtually identical to the survival benefit achieved with postoperative chemotherapy. Accordingly, neoadjuvant therapy may prove useful in selected cases (see below). A decision to use neoadjuvant chemotherapy should always be made in consultation with an experienced surgeon.

All patients with resected NSCLC are at high risk of developing a second primary lung cancer or recurrence, most of which occur within 18–24 months of surgery. Thus, it is reasonable to follow these patients with periodic imaging studies. Given the results of the NLST, periodic CT scans appear to be the most appropriate screening modality. Based on the timing of most recurrences, some guidelines recommend a contrasted chest CT scan every 6 months for the first 3 years after surgery, followed by yearly CT scans of the chest without contrast thereafter.

MANAGEMENT OF STAGE III NSCLC

Management of patients with stage III NSCLC usually requires a combined-modality approach. Patients with stage IIIB disease commonly are stratified into those with “nonbulky” or “bulky” mediastinal lymph node (N2) disease. Although the definition of “bulky” N2 disease varies somewhat in the literature, the usual criteria include the size of a dominant lymph node (i.e., >2–3 cm in short-axis diameter as measured by CT), groupings of multiple smaller lymph nodes, evidence of extracapsular nodal involvement, or involvement of more than two lymph node stations. The distinction between nonbulky and bulky stage IIIB disease is mainly used to select potential candidates for *upfront* surgical resection or for resection after neoadjuvant therapy. Many aspects of therapy of patients with stage III NSCLC remain controversial, and the optimal treatment strategy has not been clearly defined. Furthermore, because stage III disease is highly heterogeneous, no single treatment approach can be recommended for all patients. Key factors guiding treatment choices include the particular combination of tumor (T) and nodal (N) disease, the ability to achieve a complete surgical resection if indicated, and the patient’s

overall physical condition and preferences. For example, in carefully selected patients with limited stage IIIB disease where involved mediastinal lymph nodes can be completely resected, initial surgery followed by postoperative chemotherapy (with or without radiation therapy) may be indicated. By contrast, for patients with clinically evident bulky mediastinal lymph node involvement, the standard approach to treatment is concurrent chemoradiotherapy followed by a year of immunotherapy with durvalumab or other PD-L1-directed antibody.

Absent and Nonbulky Mediastinal (N2, N3) Lymph Node Disease

For the subset of stage IIIB patients initially thought to have clinical stage I or II disease (i.e., pathologic involvement of mediastinal [N2] lymph nodes is *not* detected preoperatively), surgical resection is often the treatment of choice. This is followed by adjuvant chemotherapy in patients with microscopic lymph node involvement in a resection specimen. Postoperative radiation therapy (PORT) may also have a role for those with close or positive surgical margins. Patients with tumors exceeding 7 cm in size or involving the chest wall or proximal airways within 2 cm of the carina with hilar lymph node involvement (but not N2 disease) are classified as having T3N1 stage IIIB disease. They too are best managed with surgical resection, if technically feasible, followed by adjuvant chemotherapy if completely resected. Patients with T3N0 or T3N1 disease due to the presence of satellite nodules within the same lobe as the primary tumor are also candidates for surgery, as are patients with ipsilateral nodules in another lobe and negative mediastinal nodes (IIIA, T4N0 or T4N1). Although data regarding adjuvant chemotherapy in the latter subsets of patients are limited, it is often recommended.

Patients with T4N0-1 may have involvement of the carina, superior vena cava, or a vertebral body and yet still be candidates for surgical resection in selected circumstances. The decision to proceed with an attempted resection must be made in consultation with an experienced thoracic surgeon often in association with a vascular or cardiac surgeon and an orthopedic surgeon depending on tumor location. However, if an incomplete resection is inevitable or if there is evidence of N2 involvement (stage IIIB), surgery for T4 disease is contraindicated. Most T4 lesions are best treated with concurrent chemoradiotherapy followed by durvalumab.

The role of PORT in patients with completely resected stage III NSCLC is controversial. To a large extent, the use of PORT is dictated by the presence or absence of N2 involvement and, to a lesser degree, by the biases of the treating physician. Using the Surveillance, Epidemiology, and End Results (SEER) database, a recent meta-analysis of PORT identified a significant increase in survival in patients with N2 disease but not in patients with N0 or N1 disease. An earlier analysis by the PORT Meta-analysis Trialist Group using an older database produced similar results.

Known Mediastinal (N2, N3) Lymph Node Disease When pathologic involvement of mediastinal lymph nodes is documented preoperatively, a combined-modality approach is recommended assuming the patient is a candidate for treatment with curative intent. These patients are at high risk for both local and distant recurrence if managed with resection alone. For patients with stage III disease who are not candidates for surgical resection, *concurrent* chemoradiotherapy is most commonly used as the initial treatment followed by durvalumab. Concurrent chemoradiotherapy has been shown to produce superior survival compared to *sequential* chemoradiotherapy; however, it also is associated with greater host toxicities (including fatigue, esophagitis, and neutropenia). Therefore, for patients with a good performance status, concurrent chemoradiotherapy is the preferred treatment approach, whereas sequential chemoradiotherapy may be more appropriate for patients with a performance status that is not as good. For patients who are *not* candidates for a combined-modality treatment approach, typically due to a poor performance status or a comorbidity that makes chemotherapy untenable, radiotherapy alone may provide a modest survival benefit in addition to symptom palliation.

For patients with potentially resectable N2 disease, it remains uncertain whether surgery after neoadjuvant chemoradiotherapy improves survival. In an NCI-sponsored Intergroup randomized trial comparing concurrent chemoradiotherapy alone to concurrent chemoradiotherapy followed by attempted surgical resection, no survival benefit was observed in the trimodality arm compared to the bimodality therapy. In fact, patients subjected to a pneumonectomy had a worse survival outcome. By contrast, those treated with a lobectomy appeared to have a survival advantage based on a retrospective subset analysis. Thus, in carefully selected, otherwise healthy patients with nonbulky mediastinal lymph node involvement, surgery may be a reasonable option if the primary tumor can be fully resected with a lobectomy. This is not the case if a pneumonectomy is required to achieve complete resection.

Superior Sulcus Tumors (Pancoast Tumors) Superior sulcus tumors represent a distinctive subset of stage III disease. These tumors arise in the apex of the lung and may invade the second and third ribs, the brachial plexus, the subclavian vessels, the stellate ganglion, and adjacent vertebral bodies. They also may be associated with Pancoast syndrome, characterized by pain that may arise in the shoulder or chest wall or radiate to the neck. Pain characteristically radiates to the ulnar surface of the hand. Horner's syndrome (enophthalmos, ptosis, miosis, and anhidrosis) due to invasion of the paravertebral sympathetic chain may be present as well. Patients with these tumors should undergo the same staging procedures as all patients with stage II and III NSCLC. Neoadjuvant chemotherapy or combined chemoradiotherapy followed by surgery is reserved for those without N2 involvement. This approach yields excellent survival outcomes (>50% 5-year survival in patients with an R0 resection). Patients with N2 disease are less likely to benefit from surgery and can be managed with chemoradiotherapy followed by durvalumab. Patients presenting with metastatic disease can be treated with radiation therapy (with or without chemotherapy) for symptom palliation.

MANAGEMENT OF METASTATIC NSCLC

Approximately 40% of NSCLC patients present with advanced, stage IV disease at the time of diagnosis. In addition, a significant number of patients who first presented with early-stage NSCLC will eventually relapse with distant disease. Patients who have recurrent disease have a better prognosis than those presenting with metastatic disease at the time of diagnosis. Standard medical management, the judicious use of pain medications, and the appropriate use of radiotherapy and systemic therapy—which may consist of targeted therapy, immunotherapy, and/or traditional cytotoxic chemotherapy depending on the specific diagnosis as well as PD-L1 tumor proportion score (TPS) and molecular subtype—form the cornerstone of management. Systemic therapy palliates symptoms, improves quality of life, and improves survival in patients with metastatic NSCLC, particularly in patients with good performance status. Of note, the early application of palliative care in conjunction with chemotherapy in patients with advanced NSCLC is associated with both improved survival and quality of life.

Targeted Therapies for Select Molecular Cohorts of NSCLC For a cohort of NSCLC patients, the presence of an oncogenic driver mutation allows the use of oral therapies with significant antitumor activity and improved survival compared to cytotoxic chemotherapy. These driver mutations occur in genes encoding signaling proteins that, when aberrant, promote the uncontrolled growth and metastasis of tumor cells. Importantly, driver mutations can serve as Achilles' heels for tumors, if their gene products can be targeted therapeutically with small-molecule inhibitors. All patients with advanced NSCLC should undergo molecular testing with broad panel-based testing techniques such as next-generation sequencing (NGS) to look for oncogenic drivers. Mutations, fusions, and deletions have been reported in a number of genes including *EGFR*, *ALK*, *ROS1*, *BRAF*, *RET*, *MET*, *NTRK*, *KRAS*, *PIK3CA*, *NRAS*, *AKT1*, and *MEK1* (*MAP2K1*); however, not all are considered

actionable at this time. As our treatment armamentarium expands, knowledge of these mutations is critical for selection of appropriate therapy.

EGFR mutations have been detected in 10–15% of North American patients diagnosed with NSCLC. *EGFR* mutations are associated with younger age, light (<10 pack-year) and nonsmokers, and adenocarcinoma histology. Approximately 90% of these mutations are exon 19 deletions or exon 21 L858R point mutations within the *EGFR* TK domain, resulting in hyperactivation of both *EGFR* kinase activity and downstream signaling. Lung tumors that harbor activating mutations within the *EGFR* kinase domain display high sensitivity to small-molecule *EGFR* TKIs. Osimertinib, erlotinib, gefitinib, afatinib and dacomitinib are FDA-approved oral small-molecule TKIs that inhibit *EGFR*. Several large, international, phase 3 studies have demonstrated improved response rates and progression-free survival in patients with *EGFR* mutation-positive NSCLC treated with an *EGFR* TKI as compared with standard first-line chemotherapy regimens (Table 78-9). Osimertinib was shown in a randomized phase 3 trial to have superior progression-free and overall survival in patients with *EGFR*-mutant NSCLC compared to earlier-generation *EGFR* TKIs (erlotinib or gefitinib) and to chemotherapy.

Chromosomal rearrangements involving the anaplastic lymphoma kinase (ALK) gene on chromosome 2 have been found in ~3–7% of patients with NSCLC. *ALK* rearrangements lead to hyperactivation of the *ALK* TK domain. Similar to *EGFR*, *ALK* rearrangements are typically (but not exclusively) associated with younger age, light (<10 pack-year) and nonsmokers, and adenocarcinoma histology. Crizotinib is a first-generation *ALK* TKI, whereas alectinib, brigatinib and ceritinib are second-generation *ALK* TKIs approved as first-line treatment options for patients with lung tumors harboring *ALK* rearrangements. Both alectinib and brigatinib have been found to have superior progression-free survival to crizotinib, whereas lorlatinib, a third-generation *ALK* TKI, is approved in patients who progress on a second-generation *ALK* TKI (Table 78-10). *ALK* testing may be performed via fluorescence

TABLE 78-9 Phase 3 Trials of EGFR TKIs in EGFR-Positive Non-Small-Cell Lung Cancer

TRIAL	THERAPY	NO. OF PATIENTS	ORR (%)	PFS (MONTHS)
IPASS	CbP	129	47	6.3
	Gefitinib	132	71	9.3
EURTAC	CG	87	15	5.2
	Erlotinib	86	58	9.7
OPTIMAL	CG	72	36	4.6
	Erlotinib	82	83	13.1
NEJ002	CG	114	31	5.4
	Gefitinib	114	74	10.8
WJTOG3405	CD	89	31	6.3
	Gefitinib	88	62	9.2
LUX LUNG 3	CP	115	23	6.9
	Afatinib	230	56	11.1
LUX LUNG 6	CG	122	23	5.6
	Afatinib	242	67	11.0
LUX LUNG 7	Erlotinib	159	58	10.9
	Afatinib	160	73	11.0
ARCHER 1050	Gefitinib	225	72	9.2
	Dacomitinib	227	75	14.7
FLAURA	Erlotinib or Gefitinib	277	76	8.5
	Osimertinib	279	80	17.2

Abbreviations: CbP, carboplatin and paclitaxel; CD, cisplatin and docetaxel; CG, cisplatin and gemcitabine; CP, cisplatin and paclitaxel; ORR, overall response rate; PFS, progression-free survival.

TABLE 78-10 Results of Phase 3 Trials Comparing First-Line ALK Inhibitors in ALK-Positive NSCLC

TRIAL	THERAPY	NO. OF PATIENTS	ORR (%)	PFS (MONTHS)
Profile 1014	Crizotinib	172	74	10.9
	Platinum-chemotherapy	171	45	7.0
ALEX	Alectinib	152	82.9	25.7
	Crizotinib	151	75.5	10.4
J-ALEX	Alectinib	103	92	34.1
	Crizotinib	104	79	10.2
ALTA1L	Brigatinib	137	71	67% at 1 year
	Crizotinib	138	60	43% at 1 year

Abbreviations: NSCLC, non-small-cell lung cancer; ORR, overall response rate; PFS, progression-free survival.

in situ hybridization (FISH), immunohistochemistry (IHC), or NGS.

ROS1 fusions, detected by FISH or NGS, have been identified in ~1% of patients with NSCLC, and similar to *EGFR* mutations and *ALK* fusions, *ROS1* rearrangements are typically associated with younger age and light or never smoking status. Crizotinib and lorlatinib, which inhibit both *ALK* and *ROS1* kinases, and entrectinib have been FDA approved for patients whose tumors harbor a *ROS1* fusion.

NTRK fusions occur in members of the *NTRK* gene family (*NTRK1*, *NTRK2*, *NTRK3*) and result in constitutive protein kinase activation. *NTRK* fusions are rare, occurring in <1% of patients with NSCLC. Similar to the above mutations, they more commonly occur in never smokers; however, patients with *NTRK* fusions are often older patient compared to those with *ROS1* and *ALK* alterations. Entrectinib and larotrectinib have demonstrated durable antitumor efficacy and are currently approved for *NTRK*-positive NSCLC.

MET exon 14 skipping mutations have also been identified in approximately 3–5% of patients with NSCLC. Unlike the above-mentioned mutations, *MET* exon 14 skipping mutations occur in both squamous and nonsquamous NSCLC patients and those with a history of smoking. Pharmacologic inhibition of the overactive *MET* pathway with capmatinib or tepotinib resulted response rates >70%, particularly in treatment-naïve NSCLC patients.

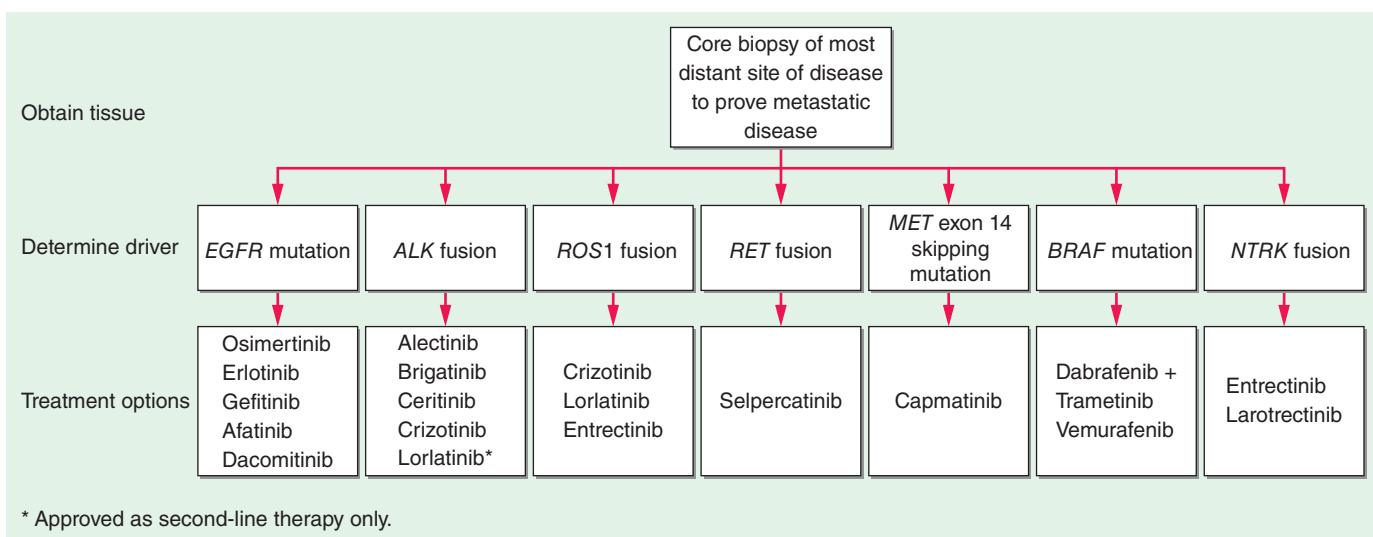
Oncogenic mutations in *BRAF* have been observed in ~2% of patients with NSCLC and, similar to *MET*, occur in both squamous and nonsquamous NSCLC and with an equal prevalence in patients

with a history of smoking. This mutation is typically most targetable when it occurs at the 600th amino acid valine (V600). Combined inhibition with a *BRAF* and MEK inhibitor, dabrafenib plus trametinib, is a first-line or later therapeutic option in patients with *BRAF* V600-mutant NSCLC and appears to be superior to *BRAF* or MEK inhibition alone.

RET alterations typically occur as chromosomal rearrangements resulting in constitutive TKI activation. *RET* rearrangements may be detected by either FISH or NGS in ~1% of NSCLC patients. Analogous to capmatinib, selpercatinib has demonstrated an excellent response rate; as many as 85% of treatment-naïve NSCLC patients with *RET* alterations responded. All National Comprehensive Cancer Network-supported targetable oncogenic driver mutations and potential therapeutic options are summarized in Fig. 78-7.

Mutations within the *KRAS* GTPase are found in ~20% of lung adenocarcinomas. Agents targeting *KRAS* G12C are in development. Each of the other driver mutations occurs in <1–3% of lung adenocarcinomas. The great majority of the driver mutations are mutually exclusive. Most cancers have just one main driver. Defining mechanisms of acquired resistance to small-molecule inhibitors is a high research priority.

Immunotherapy Immune checkpoint inhibitors have significantly improved the quality of life and survival for a group of patients with locally advanced and metastatic NSCLC. These agents are used primarily in patients whose tumors do not express a targetable genetic lesion (Fig. 78-8). Immune checkpoint inhibitors work by blocking interactions between T cells and antigen presenting cells (APCs) or tumor cells that lead to T-cell inactivation. By inhibiting this interaction, the immune system is effectively upregulated and T cells become activated against tumor cells. Several randomized studies have demonstrated superior overall survival in patients treated with pembrolizumab or atezolizumab monotherapy or nivolumab plus ipilimumab combination immunotherapy compared to chemotherapy in patients with metastatic NSCLC with PD-L1 expression in ≥50% of tumor cells (Keynote 024, IMPOWER 110) and ≥1% of tumor cells (Keynote 042, CheckMate 227) (Table 78-11). The evidence supporting the use of single-agent immunotherapy in patients with tumor PD-L1 <50% remains unclear; current recommendations suggest the use of chemotherapy plus immunotherapy or immunotherapy combinations as the first-line treatment strategy in patients with metastatic NSCLC with tumor PD-L1 <50%. As discussed below, specific regimens vary by tumor histology (adenocarcinoma vs squamous cell carcinoma). Although PD-L1 has been identified as a biomarker that can predict response to immune checkpoint inhibitors, responses are observed in patients who do

**FIGURE 78-7 Approach to targeted therapy in non-small-cell lung cancer (NSCLC).**

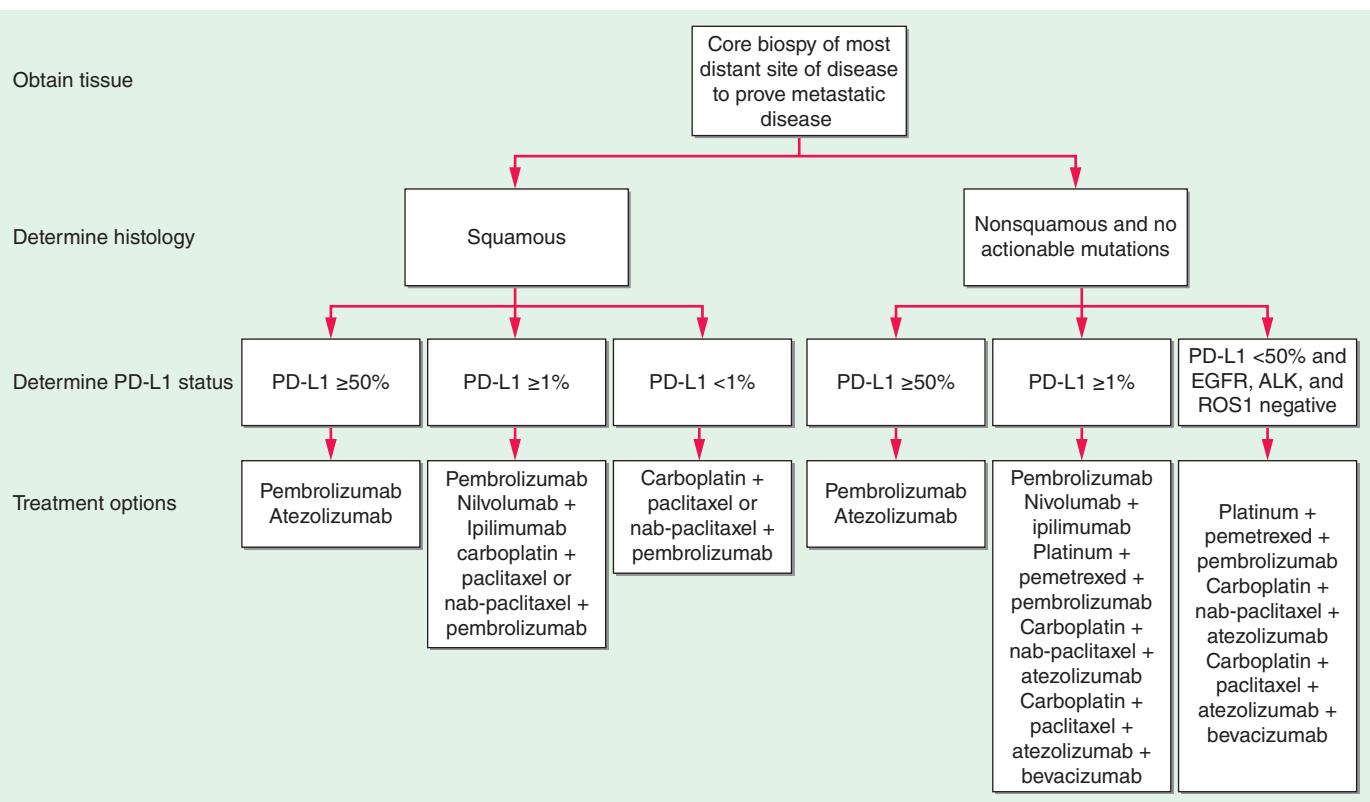


FIGURE 78-8 Approach to first-line therapy in a patient with stage IV, driver mutation-negative non-small-cell lung cancer (NSCLC).

not appear to express the biomarker, and not all PD-L1-positive patients respond to checkpoint inhibition. Importantly patients with driver mutations such as *EGFR* and *ALK* appear to derive greater benefit from targeted therapy than immunotherapy and should be treated with a TKI, even in the presence of high PD-L1 expression. Further evaluation of these agents in the neoadjuvant setting and combined with chemoradiotherapy is ongoing.

Cytotoxic Chemotherapy for Metastatic or Recurrent NSCLC

Cytotoxic chemotherapy is typically used in combination with immunotherapy as the initial treatment in patients with metastatic or recurrent NSCLC only when there is no contraindication to immunotherapy. Selected chemotherapy agents perform quite differently in squamous carcinomas versus adenocarcinomas. Patients with nonsquamous NSCLC have an improved survival when

TABLE 78-11 Results of Phase 3 Trials Comparing First-Line Immunotherapy with or without Chemotherapy Versus Chemotherapy Alone in Patients with NSCLC

STUDY	THERAPY	NO. OF PATIENTS	OS (MONTHS)	PFS (MONTHS)
Keynote 024 PD-L1 ≥50%	Pembrolizumab	154	30.0	7.9
	Platinum-chemotherapy	151	14.2	3.5
Keynote 042 PD-L1 ≥1%	Pembrolizumab	637	16.7	5.4
	Platinum-chemotherapy	637	12.1	6.5
IMPOWER 110 PD-L1 ≥50% TC or ≥15% IC	Atezolizumab	286	20.2	8.1
	Platinum-chemotherapy	263	13.1	5.0
Keynote 189 Nonsquamous	Pembrolizumab + platinum-chemotherapy	410	NR	8.8
	Platinum-chemotherapy	206	11.3	4.9
Keynote 407 Squamous	Pembrolizumab + platinum-chemotherapy	278	15.9	6.4
	Platinum-chemotherapy	281	11.3	4.8
IMPOWER 150 Nonsquamous	Atezolizumab + platinum-chemotherapy	356	19.2	8.3
	Platinum-chemotherapy	336	14.7	6.8
IMPOWER 130 Nonsquamous	Atezolizumab + platinum-chemotherapy	483	18.6	7.0
	Platinum-chemotherapy	240	13.9	5.5
CheckMate 227	Nivolumab plus ipilimumab	583	17.1	5.1
	Platinum-chemotherapy	583	13.9	5.6
CheckMate-9LA	Nivolumab plus ipilimumab plus two cycles of platinum-chemotherapy	361	14.1	6.8
	Platinum-chemotherapy	358	10.7	5

Abbreviations: IC, immune cells; NR, not reported; OS, overall survival; PFS, progression-free survival; TC, tumor cells.

Note: Platinum-chemotherapy refers to first-line platinum doublet or triplet chemotherapy.

treated with cisplatin and pemetrexed compared to cisplatin and gemcitabine. By contrast, patients with squamous carcinoma have an improved survival when treated with cisplatin and gemcitabine. This survival difference is thought to be related to the differential expression between tumor types of thymidylate synthase (TS). Squamous cancers have a much higher expression of TS compared to adenocarcinomas, accounting for their lower responsiveness to pemetrexed. By contrast, the activity of gemcitabine is not impacted by the levels of TS.

Second-Line Therapy and Beyond Second-line therapy for advanced NSCLC relies on docetaxel; it improves survival compared to supportive care alone. Ramucirumab is a recombinant human IgG1 monoclonal antibody that targets VEGFR-2 and blocks the interaction of VEGF ligands and VEGFR-2. A phase 3 trial demonstrated a significant improvement in progression-free survival and overall survival when ramucirumab was combined with docetaxel as second-line therapy in patients who had progressed on platinum-based chemotherapy. Contrary to bevacizumab, ramucirumab was safe in patients with both squamous and nonsquamous NSCLC and is approved regardless of histology.

Supportive Care No discussion of the treatment strategies for patients with advanced lung cancer would be complete without a mention of supportive care. Coincident with advances in chemotherapy and targeted therapy was a pivotal study that demonstrated that the early integration of palliative care with standard treatment strategies improves both quality of life and overall survival for patients with stage IV NSCLC (Chaps. 12 and 69). Aggressive pain and symptom control are important components of optimal treatment of these patients.

TREATMENT

Small-Cell Lung Cancer

The overall treatment approach to patients with SCLC is shown in Fig. 78-9.

SURGERY FOR LIMITED-DISEASE SMALL-CELL LUNG CANCER

SCLC is a highly aggressive disease characterized by its rapid doubling time, high growth fraction, early development of disseminated disease, and dramatic response to first-line chemotherapy and radiation. In general, surgical resection is *not* routinely recommended for patients because even patients with LD-SCLC still have occult micrometastases. However, the American College of Chest Physicians Evidence-Based Clinical Practice Guidelines recommend surgical resection over nonsurgical treatment in SCLC patients with clinical stage I disease after a thorough evaluation for distant metastases and invasive mediastinal stage evaluation (grade 2C). After resection, these patients should receive platinum-based adjuvant chemotherapy (grade 1C). If the histologic diagnosis of SCLC is made in patients on review of a resected surgical specimen, such patients should receive standard SCLC chemotherapy as well.

CHEMOTHERAPY

In patients with limited-stage SCLC, concurrent chemoradiotherapy with cisplatin-etoposide for four cycles has remained standard of care for over 4 decades. Two randomized phase 3 trials have demonstrated that chemotherapy with either cisplatin or carboplatin plus either etoposide and a PD-L1 inhibitor, atezolizumab (IMPOWER 133) or durvalumab (CASPIAN), provides superior progression-free and overall survival compared to chemotherapy alone, making combination therapy the preferred choice in appropriate patients. Despite response rates to first-line therapy as high as 80%, the median survival ranges from 12 to 20 months for patients with LD and approximately 12 months for patients with ED. Regardless of disease extent, the majority of patients relapse and develop chemotherapy-resistant disease. The prognosis is especially poor for patients who relapse within the first 3 months of therapy; these patients are said to have *chemotherapy-resistant disease*. Patients are said to have *sensitive disease* if they relapse >3 months after their initial therapy and are thought to have a somewhat better overall survival. These patients also are thought to have the greatest potential benefit from second-line chemotherapy. Topotecan and

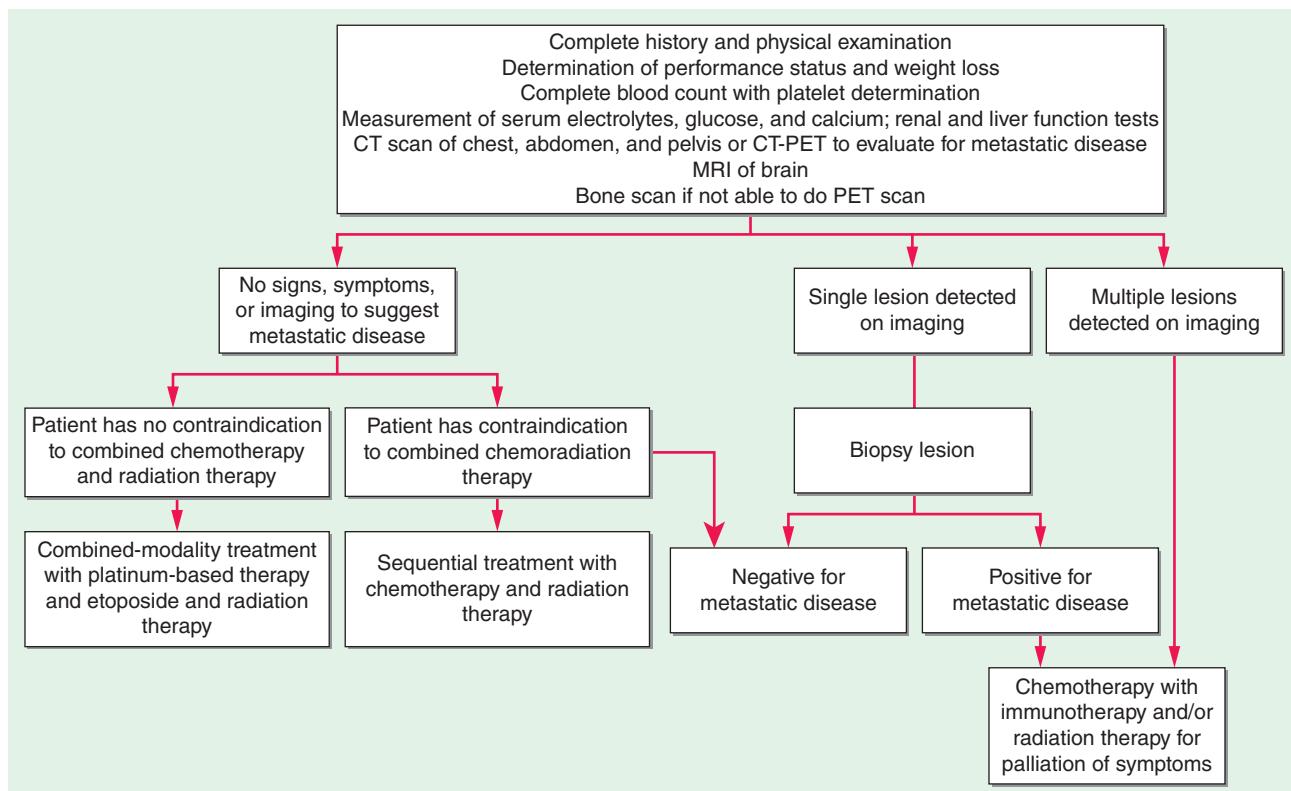


FIGURE 78-9 Algorithm for management of small-cell lung cancer. MRI, magnetic resonance imaging; PET, positron emission tomography.

Lurbinectedin are FDA-approved agents for second-line therapy in patients with SCLC. Topotecan has only modest activity and can be given either intravenously or orally; it appears to have more efficacy in patients with chemotherapy-sensitive disease. Lurbinectedin has a 35% response rate and progression-free survival of 3.5 months, with a greater benefit in patients with chemotherapy-sensitive disease. Other agents with similar low levels of activity in the second-line setting include irinotecan, paclitaxel, docetaxel, vinorelbine, oral etoposide, and gemcitabine.

THORACIC RADIATION THERAPY

Thoracic radiation therapy (TRT) is a standard component of induction therapy for patients with good performance status and limited-stage SCLC. Meta-analyses indicate that chemotherapy combined with chest irradiation improves 3-year survival by ~5% as compared with chemotherapy alone. The 5-year survival rate, however, remains disappointingly low at ~10–15%. Most commonly, TRT is combined with cisplatin and etoposide chemotherapy due to a superior toxicity profile as compared to anthracycline-containing chemotherapy regimens. As observed in locally advanced NSCLC, *concurrent* chemoradiotherapy is more effective than *sequential* chemoradiation but is associated with significantly more esophagitis and hematologic toxicity. Ideally TRT should be administered with the first two cycles of chemotherapy because later application appears slightly less effective. If for reasons of fitness or availability, this regimen cannot be offered, TRT should follow induction chemotherapy. With respect to fractionation of TRT, twice-daily 1.5-Gy fractionated radiation therapy has been shown to improve survival in LD-SCLC patients but is associated with higher rates of grade 3 esophagitis and pulmonary toxicity. Although it is feasible to deliver once-daily radiation therapy doses up to 70 Gy concurrently with cisplatin-based chemotherapy, there are no data to support equivalency of this approach compared with the 45-Gy twice-daily radiotherapy dose. Therefore, the current standard regimen of a 45-Gy dose administered in 1.5-Gy fractions twice daily for 30 days is being compared with higher-dose regimens in two phase 3 trials, one in the United States and one in Europe. Patients should be carefully selected for concurrent chemoradiation therapy based on good performance status and adequate pulmonary reserve. The role of radiotherapy in ED-SCLC is largely restricted to palliation of tumor-related symptoms such as bone pain and bronchial obstruction.

PROPHYLACTIC CRANIAL IRRADIATION

Prophylactic cranial irradiation (PCI) should be considered in all patients either with LD-SCLC or who have responded well to initial therapy; its role in patients with ED-SCLC is more controversial. A meta-analysis including seven trials and 987 patients with LD-SCLC who had achieved a complete remission after upfront chemotherapy yielded a 5.4% improvement in overall survival for patients treated with PCI. In patients with ED-SCLC who have responded to first-line chemotherapy and had no CNS disease, patients randomized to observation had a higher incidence of brain metastases; however, use of PCI did not improve overall survival. Long-term toxicities, including deficits in cognition, have been reported after PCI but are difficult to sort out from the effects of chemotherapy or normal aging.

THYMIC TUMORS

Thymic tumors are rare malignancies accounting for 0.5–1.5% of all malignancies in the United States with a higher incidence among Asian populations. They are particularly rare among children and young adults with incidence peaking in the fifth decade of life. There is no difference between sexes, and no clear risk factors have been identified.

CLINICAL MANIFESTATIONS

The majority of thymic tumors occur in the anterior mediastinum. Approximately 40% of patients with mediastinal masses will be

asymptomatic with an incidental finding on chest imaging. In patients presenting with an anterior mediastinal mass, if appropriate, serum β -human chorionic gonadotropin (HCG) and α fetoprotein (AFP) should be sent to rule out a germ cell tumor. A patient with a sign or symptom of thymoma or thymic carcinoma may present with chest pain, dyspnea, cough, or superior vena cava syndrome secondary to effects on adjacent organs or a paraneoplastic syndrome, most commonly myasthenia gravis, pure red cell aplasia, or hypogammaglobulinemia. More rare paraneoplastic syndromes include limbic encephalitis, aplastic anemia, hemolytic anemia, and autoimmune disease such as Sjögren's syndrome, polymyositis, rheumatoid arthritis, and ulcerative colitis, among others.

STAGING

Given the rarity of the tumor, patients with suspected thymoma should be evaluated by a multidisciplinary team including a surgeon, medical and radiation oncologist, and pathologist with experience in treating the disease. A CT scan of the chest with contrast is recommended to determine if the mass is resectable based on relationship to surrounding structures. An MRI with contrast may be performed if clinically indicated. A PET scan may be useful in the evaluation of a patient with thymic tumors, although it may be less useful in the staging of thymoma compared to thymic carcinoma. A core needle biopsy is considered standard of care for obtaining a histologic diagnosis of an anterior mediastinal tumor. This may be obtained via CT or ultrasound imaging. However, in some circumstances, a mediastinoscopy or open biopsy may be required.

Thymomas are commonly staged using the Masaoka system or the World Health Organization (WHO) staging system, as described in **Table 78-12**. WHO types A, AB, and B1 tend to be more well-differentiated, types B2 and B3 are moderately differentiated, and type C is poorly differentiated.

TREATMENT

Surgical resection is the mainstay of treatment for patients with Masaoka type I and II thymic tumors. In patients with type III and IV who have potentially resectable thymic tumors, neoadjuvant chemotherapy may be given to decrease the tumor size and allow for a resection with negative margins. Surgery remains controversial and provides a limited role in the treatment of stage III and IV disease. No

TABLE 78-12 Staging Thymic Tumors

MASAOKA STAGE	DEFINITION
I	Grossly and microscopically encapsulated
IIA	Microscopic transcapsular invasion
IIB	Macroscopic invasion into surrounding tissue excluding pericardium, lung, and great vessels
III	Macroscopic invasion into neighboring organs of the lower neck or upper chest
IVA	Pleural or pericardial dissemination
IVB	Hematogenous or lymphatic dissemination to distal organs

WHO	
A	Tumor with few lymphocytes
AB	Tumor with features of type A and foci rich in lymphocytes
B1	Tumor with features of normal epithelial cells with vesicular nuclei and distinct nucleoli and an abundant population of lymphocytes. Also known as cortical thymoma, lymphocyte-rich thymoma
B2	Thymoma with no or mild atypia with round or polygonal-shaped cells with small component of lymphocytes
B3	Well-differentiated thymic carcinoma with mild atypia
C	Thymic carcinoma with high atypia

additional therapy may be required in patients with type I who have a resection with negative margins. Postoperative radiation therapy may be recommended based on extracapsular extension and the presence of positive margins in patients with type II or III thymic tumors or histologic evaluation WHO B3 and C. Radiation therapy may be beneficial in patients with locally advanced disease (type III or IV) or in patients with symptoms secondary to compression of surrounding structures. Chemotherapy with cisplatin, doxorubicin, and cyclophosphamide (CAP) remains the mainstay of therapy in the neoadjuvant and adjuvant setting as well as first-line therapy in patients with metastatic thymoma, whereas carboplatin and paclitaxel are often employed in patients with thymic carcinoma. Limited additional agents are recommended based on small phase 2 trials as second-line therapy and beyond.

COVID-19 AND LUNG CANCER

COVID-19, a respiratory tract infection caused by SARS-CoV-2, emerged in Wuhan, China, in late 2019. The rapid global spread led the WHO to declare a pandemic in early March 2020. Large retrospective data sets have shown that cancer patients, and particularly patients with lung cancer, are at increased risk of morbidity and mortality from COVID-19. The dilemma of distinguishing COVID-19 symptoms from lung cancer and radiographic diagnosis of pneumonia or pneumonitis from radiation therapy or immunotherapy versus COVID-19 pneumonia has presented a particular challenge to health care providers. Mortality as high as 35% has been reported for patients with lung cancer infected with SARS-CoV-2. Older patients (≥ 65 years old), patients with a worse performance status (Eastern Cooperative Oncology Group performance status ≥ 1), patients on glucocorticoids (≥ 10 mg prednisone equivalent) and anticoagulation, and patients on chemotherapy within 3 months of diagnosis appear to be particularly at risk for mortality if infected. The long-term impact on lung cancer mortality due to delays in screening, diagnosis, and treatment are likely to be felt for years to come.

SUMMARY

The management of SCLC and NSCLC has undergone major change in the past decade, resulting in a reduction in lung cancer mortality. For patients with early-stage disease, advances in radiotherapy and surgical procedures as well as new systemic therapies have greatly improved prognosis in all diseases. For patients with advanced lung cancer, major progress in understanding tumor genetics and tumor immunology has led to the development of rational targets and specific inhibitors, which have documented efficacy in specific subsets of NSCLC. Furthermore, increased understanding of how to activate the immune system to drive antitumor immunity has proven to be a successful therapeutic strategy for a subset of patients with advanced lung cancer. However, only a small subset of patients responds to immune checkpoint inhibitors, and the majority of patients treated with targeted therapies or chemotherapy eventually develop resistance, which provides strong motivation for further research and enrollment of patients onto clinical trials in this rapidly evolving area.

ACKNOWLEDGMENT

David Johnson and Christine Lovly contributed to this chapter in the prior edition, and material from that chapter has been retained here.

FURTHER READING

- DRILLON et al: Selpercatinib. Available at <https://www.nejm.org/doi/full/10.1056/NEJMoa2005653>.
- DRILON et al: Entrectinib. Available at [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(19\)30690-4/fulltext](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(19)30690-4/fulltext)
- DRILON et al: Larotrectinib. Available at <https://www.nejm.org/doi/full/10.1056/NEJMoa1714448>.
- GHANDI L et al: Pembrolizumab plus chemotherapy in metastatic non-small cell lung cancer. N Engl J Med 378:2078, 2018.

GOLDSTRAW et al: The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. JTO 11:39, 2016.

HELLMANN MD et al: Nivolumab plus ipilimumab in advanced non small cell lung cancer. N Engl J Med 381:2220, 2019.

GARASINO et al: COVID. Available at <https://www.sciencedirect.com/science/article/pii/S1470204520303144?via%3Dhub>.

HORN L et al: First line atezolizumab plus chemotherapy in extensive stage small-cell lung cancer. N Engl J Med 379:2220, 2018.

LI et al: Trastuzumab-Deruxtecan. Available at <https://www.nejm.org/doi/full/10.1056/NEJMoa2005653>.

PAZ-ARES L et al: Pembrolizumab plus chemotherapy for squamous non-small cell lung cancer. N Engl J Med 379:2040, 2018.

PAZ-ARES L et al: Durvalumab plus platinum-etoposide versus platinum topoisomerase in first line treatment of extensive-stage small-cell lung cancer (CASPIAN): A randomised, controlled, open-label, phase 3 trial. Lancet 394:1929, 2019.

PETERS S et al: Alectinib versus crizotinib in untreated ALK-positive non-small cell lung cancer. N Engl J Med 377:829, 2017.

RECK M et al: Pembrolizumab versus chemotherapy for PD-L1 positive non-small cell lung cancer. N Engl J Med 375:1823, 2016.

SORIA JC et al: Osimertinib in untreated EGFR-mutated advanced nonsmall cell lung cancer. N Engl J Med 378:113, 2018.

WOLF et al: Capmatinib. Available at <https://www.nejm.org/doi/full/10.1056/NEJMoa2005653>.

79

Breast Cancer

Daniel F. Hayes, Marc E. Lippman

INTRODUCTION AND BACKGROUND

CONCEPTUAL AND BIOLOGICAL ISSUES OF BREAST CANCER

Breast cancer is a malignant proliferation of epithelial cells lining the ducts or lobules of the breast. In the year 2020, approximately a quarter million cases of invasive and 61,000 cases of *in situ* breast cancer were diagnosed in the United States, with nearly 41,000 deaths. Epithelial malignancies of the breast are the most common cause of cancer in women (excluding skin cancer), accounting for about one-third of all cancer in women. As a result of earlier detection and improved treatments, the mortality rate from breast cancer decreased by more than one-third over the past three decades in high- and middle-income countries. This chapter does not consider rare malignancies presenting in the breast, such as sarcomas and lymphomas, but focuses on the epithelial cancers.

Breast cancer has served as a paradigm for several oncologic principles related to solid tumors. It spans a spectrum of conditions for which different clinical considerations must be made, including risk assessment, prevention, screening, evaluation of breast abnormalities, local and adjuvant systemic treatments, metastatic therapies, and survivorship issues (Fig. 79-1).

The unique biology of breast cancer has rendered it amenable to a variety of therapeutic “targeted” strategies based on the appreciation of differences in subtypes that reflect the need for differences in assessment and therapy. These subtypes include expression of the estrogen receptor (ER) and the human epidermal growth factor receptor type 2 (HER2), as well as germline or somatic mutations in inherited tumor suppressor genes, such as *BRCA1* and *BRCA2*. Identifiable somatic mutations in genes that appear to drive the cancer, including mammalian target of rapamycin (*mTOR*), cyclin-dependent kinase 4 and 6 (*CDK4/6*), and S-phosphatidylinositol-4,5-bisphosphate 3-kinase

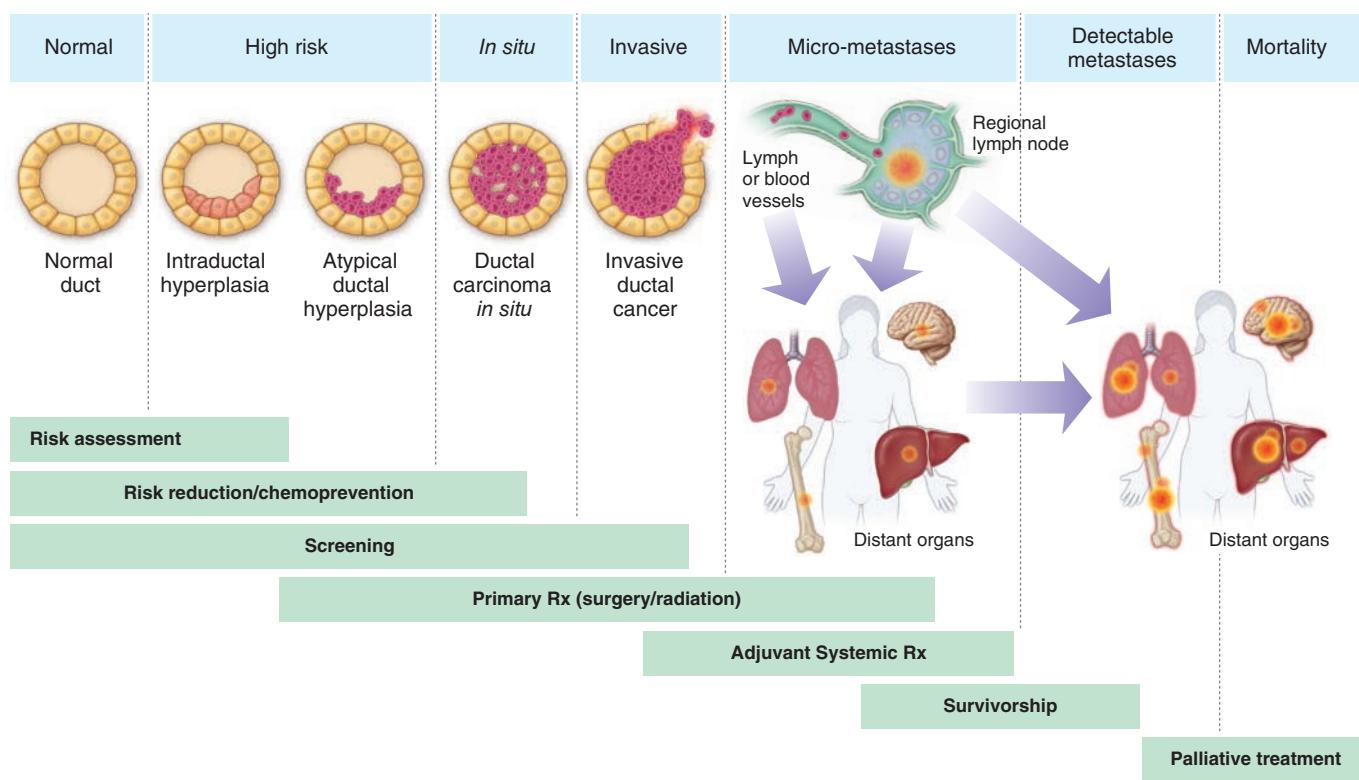


FIGURE 79-1 Breast cancer continuum conceptual model. Most breast cancers begin in epithelial cells within the lobules or ducts. They proceed through a continuum of atypia and hyperplasia to *in situ* malignancy to invasion into surrounding normal tissues followed by intravasation into lymph and blood channels to local lymph nodes and distant organs, culminating in distant metastases. This is a conceptual model. Not all metastatic breast cancers have progressed through these stages, and many lesions do not progress to the next.

catalytic subunit alpha (*PIK3CA*) make it susceptible to specific therapeutic interventions directed against each of these targets (Table 79-1). Furthermore, immune checkpoint inhibition has been applied to specific types of breast cancers.

EPIDEMIOLOGY AND RISK FACTORS

■ CLINICAL, HORMONAL, AND OTHER NONGENETIC RISK FACTORS

Seventy-five percent of all breast cancers occur in women aged >50 years, but breast cancer is not uncommon in women in their forties and can occur in women in their thirties and even twenties, and very rarely in adolescence.

Breast cancer is principally a sex hormone-dependent disease through increased activity of the ER and its ligands, estradiol and estrone (Fig. 79-1, Table 79-1). Indeed, the female-to-male ratio is ~150:1. Relative exposure to both endogenous and exogenous estrogens increases risk of breast cancer. Risk of developing breast cancer is higher in women with early menarche (<12 years) and late first full-term pregnancy (>35 years), and it is increased by exogenous hormone replacement therapy. Women without functioning ovaries, who experience an early menopause, or who never receive combination estrogen/progesterone replacement therapy are much less likely to develop breast cancer than those who have a normal menstrual history. Also, duration of maternal nursing correlates with substantial risk reduction independent of either parity or age at first full-term pregnancy.

Menstrual and reproductive history accounts for 70–80% of the variation in breast cancer frequency in different countries, providing insight into hormonal carcinogenesis. A woman living to age 80 years in North America has a one in nine chance of developing invasive breast cancer. Women who live in agrarian societies, especially in Asia, have traditionally had only 1/5th to 1/10th the risk of breast cancer of

women in North America or Western Europe. However, Asian women who immigrate to North America or European countries during preadolescence or in adolescence have the same risk as women born in these countries. Further, with shifts from agrarian to industrialized economic systems, the incidence of breast cancer has risen dramatically in Asia, approaching that observed in Western nations.

Exogenous use of female hormones also plays a role in breast cancer incidence. The elevated risk related to oral contraceptives is quite modest if present at all. Regardless, this risk is more than balanced by avoidance of an undesired pregnancy and a substantial protective effect against ovarian epithelial and endometrial cancers.

Hormone replacement therapy (HRT) with conjugated equine estrogens plus progestins increases the risk of breast cancer; 6–7 years of HRT nearly doubles the risk of breast cancer and also increases the incidence of adverse cardiovascular events. However, it decreases the risk of bone fractures and colorectal cancer. On balance, more negative than positive events are associated with HRT. Administration of conjugated estrogens is usually combined with companion progesterone to abrogate the increased risk of endometrial cancer with estrogen alone. However, single-agent estrogen replacement therapy in women who have had hysterectomies produces no significant increase in breast cancer incidence and, if anything, reduces the risk. Thus, there are serious concerns about long-term HRT, especially in combination with progestins, in terms of cardiovascular disease and breast cancer. No comparable safety data are available for other less potent forms of estrogen replacement, such as bioequivalent estrogen found in soy, and they should not be routinely used as substitutes. Epidemiologic studies demonstrate a rapid decrease in elevated breast cancer incidence coincident with discontinuation of HRT.

HRT in women previously diagnosed with breast cancer, especially of the subtype that expresses ERs, counteracts much of the effectiveness of antineoplastic endocrine therapies and is contraindicated. Although

TABLE 79-1 Breast Cancer Molecular Features and Associated Targeted Therapies

MOLECULE	GENE THAT ENCODES MOLECULE	ABNORMALITY	CLASS OF TARGETED THERAPIES	SPECIFIC THERAPIES
Estrogen receptor (ER)	<i>ESR1</i>	Overexpression of cellular protein	Endocrine therapies	
			Estrogen ablation (surgical, chemical)	Premenopausal Oophorectomy Luteinizing hormone releasing hormone (LHRH) agonists (goserelin, leuprorelin) or antagonist (triptorelin) Postmenopausal Aromatase inhibitors (AIs) (anastrozole, letrozole, exemestane)
			ER antagonists	Selective estrogen receptor modulators (SERMs) (tamoxifen, toremifene, raloxifene) Selective estrogen receptor downregulators (SERDs) (fulvestrant)
Human epidermal growth factor receptor type 2 (HER2)	<i>c-neu/erbB2</i>	Overexpression of cell surface protein	Antibodies against HER2	Trastuzumab, pertuzumab, margetuximab
			Antibody-drug conjugates against HER2	Ado-trastuzumab emtansine, fam-trastuzumab deruxtecan-nxki
			Tyrosine kinase inhibitors	Lapatinib, neratinib, tucatinib
		Mutations	Tyrosine kinase inhibitors	Neratinib (<i>indication not FDA approved</i>)
Mammalian target of rapamycin (mTOR)	<i>MTOR</i>	Loss of protein suppressor of mTOR pathway, phosphatase and tensin homolog (<i>PTEN</i>)	Tyrosine kinase inhibitor	Everolimus
Cyclin-dependent kinase 4 and 6 (CDK4/6)	<i>CDK4, CDK6</i>	Loss of the protein suppressor of CDK4/6 pathway, retinoblastoma (<i>RB1</i>)	Inhibition of CDK4/6 enzyme activity	Palbociclib, ribociclib, abemaciclib
Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA)	<i>PIK3CA</i>	Mutations	Enzyme inhibition of mutated/activated PIK3CA protein	Alpelisib
BRCA1/2	<i>BRCA1, BRCA2</i>	Loss of tumor suppressor activity of BRCA1/2	Inhibition of poly (ADP-ribose) polymerase (PARP) activity and synthetic lethality	PARP inhibitors (olaparib, talazoparib)
TROP-2	<i>TACSTD2</i>	Over expression of TROP-2 cell surface protein	Antibody-drug conjugate against TROP-2	Sacituzumab govitecan
Immune checkpoints	NA	Programmed death-ligand 1 (PD-L1) suppression of immune effector cells	Inhibition of PD-L1/PD-1 suppression of immune effector cells	Atezolizumab

Abbreviations: FDA, U.S. Food and Drug Administration; NA, not applicable.

intravaginal estrogen therapy has been used for atrophic vaginitis associated with antiestrogenic endocrine therapies, it does result in some absorption and systemic estrogenic effects and should generally be avoided.

In addition to sex, age, and hormonal exposure, other risk features for breast cancer have been identified, but none with the kind of relative, attributable, or absolute risks associated with these three factors. Various differences in diets (including Asian agrarian vs modern economic) have been implicated as risk factors, although the role of diet in breast cancer etiology is controversial. Associative links exist between breast cancer risk and total caloric and fat intake, or even specific types of caloric intake, but the exact role of fat in the diet is unproven and may actually intersect with menstrual history and estrogenic exposure.

Central obesity, metabolic syndrome, and type 2 diabetes mellitus are all risk factors for occurrence and recurrence of breast cancer. Moderate alcohol intake also increases the risk by an unknown mechanism. Folic acid supplementation appears to modify risk in women who use alcohol but is not additionally protective in abstainers. Recommendations favoring abstinence from alcohol must be weighed against other social pressures and the possible cardioprotective effect of moderate alcohol intake. Depression is also associated with both occurrence and recurrence of breast cancer.

Certain benign breast pathologic findings, such as atypical hyperplasia and radial scars, are associated with higher risk of subsequent breast cancers. Prior radiation is a risk factor, but principally when delivered in adolescence or early child-bearing ages. Women who have been exposed before age 30 years to radiation in the form of multiple fluoroscopies (200–300 cGy) or treatment for Hodgkin's disease (>3600 cGy) have a substantial increase in risk of breast cancer, whereas radiation exposure after age 30 years appears to have a minimal carcinogenic effect on the breast. Radioactive iodine therapy for thyroid disease is not associated with increased risk of breast cancer, whereas mediastinal radiation in younger women for lymphoma is a powerful risk factor within the radiated field.

■ INHERITED GERMLINE SUSCEPTIBILITY FACTORS

 Family history has long been recognized as a risk factor for breast cancer. A woman with a first-degree history (mother or sister) of breast cancer has an increased relative risk of approximately 30–50% (or one-third to one-half higher) over a woman with no family history. However, family history only accounts for 10–15% of all breast cancers. Most women who develop breast cancer do not have a strong family history. For women without an identifiable inherited genetic

abnormality, it is not clear whether the increased risk associated with family history is due to environmental causes or as yet unidentified genetic abnormalities.

The genetics of breast cancer require an understanding of the distinction between inherited, germline genetic differences among individuals and acquired, somatic genetic changes within cancers. The former are often called mutations but are more properly termed *single nucleotide polymorphisms* (SNPs). Some SNPs are synonymous, meaning they do not change the encoded amino acid in the affected protein product, and therefore are of no clinical significance. Some SNPs are nonsynonymous but may lead to a substituted amino acid that does not change the function of the protein, and they are likewise clinically insignificant. However, if an SNP leads to an amino acid substitution that alters the protein function or results in complete cessation of transcription or translation (a “stop” codon), it is considered deleterious and leads to higher susceptibility to developing cancer. In some cases, the significance of the SNP is unknown, and these are designated *variants of undetermined significance* (VUS).

The genes of interest serve, in the normal cell, to suppress expansion of a potentially malignant clone, either by repairing downstream randomly occurring somatic genetic abnormalities or, if not possible, by inducing programmed cell death, or apoptosis. Somatic genetic changes that are not inherited, including mutations, amplifications, insertions, deletions, translocations, and others, are responsible for the malignant behavior of a cancer, including unrestrained proliferation, as well as extravasation from one site and migration and establishment of metastases into another. As discussed below, some germline and somatic mutations can be exploited therapeutically (Table 79-1).

For most women, the increased risk associated with a family member who has had breast cancer appears to be related to both a weak, and probably multigene, germline susceptibility and similar exposure to environmental/lifestyle risk factors. Only approximately 10% of human breast cancers can be linked directly to a single inherited germline SNP. However, when one is present, the relative and absolute risks for that individual developing breast, and other, cancers in her lifetime are extraordinary.

The *BRCA1* and *BRCA2* genes, located on chromosomal loci 17q21 and 13q12, respectively, are the best characterized breast cancer susceptibility genes and have the greatest clinical importance in assessing genetic risk for breast cancer. Women who inherit mutated alleles of these genes from either parent have at least a 60–80% lifetime chance of developing breast cancer and about a 33% chance of developing ovarian cancer. The cancers that arise within a *BRCA1*-mutated patient are almost exclusively negative for ER, progesterone receptor (PgR), and HER2 (so-called “triple-negative” breast cancers). Men who carry a mutant allele of the gene have an increased incidence of breast and also prostate cancers, although the absolute risk of breast cancer in men with *BRCA2* germline SNPs is far lower than that for women who harbor them.

Overall, <1% of the general population and <5% of all patients with breast cancer harbor deleterious SNPs in *BRCA1* or *BRCA2*. Certain subgroups of women are more likely to have *BRCA1/2* mutations. The incidence is approximately 2% in women of Ashkenazi, Eastern European descent. Approximately 20% of women with triple-negative breast cancers will be positive for deleterious germline *BRCA1* SNPs, and genetic testing is warranted in most patients with triple-negative breast cancer even without a family history.

In contrast to those that arise in *BRCA1* carriers, cancers that arise in *BRCA2* contexts are more likely to be ER positive. The incidence of *BRCA2* mutations is much higher than *BRCA1* in men who develop breast cancer. However, most male breast cancer cases do not occur in *BRCA1/2*-mutated men, and the risk of breast cancer in men who do carry the *BRCA2* mutation is much lower than that in women with this genetic abnormality. Many other inherited germline mutations in known or putative tumor-suppressor genes, such as *p53* (which also accounts in part for the Li-Fraumeni familial cancer syndrome), *PTEN* (which accounts for Cowden’s syndrome), and *PALB1*, have now been identified as important tumor-suppressor genes with clinical implications.

Inherited germline mutations are readily detected in blood tests of normal circulating leukocytes using so-called “panel” assays, which at present provide results from 30–45 different germline genes. However, because the rate of deleterious germline SNPs in these genes in the general population is quite low (well below 1%), germline panel genetic testing of the entire population of women is not recommended. Further, results are confounded by the presence of VUS in known tumor-suppressor genes, such as *BRCA1* and *BRCA2*, or deleterious variants or VUS in genes that are putative, but not proven, tumor-suppressor genes. Such results lead to confusion, anxiety, and inappropriate preventive strategies, such as prophylactic surgery, in individuals who may not actually be at higher risk.

Consensus guidelines on who should be tested include any woman with a family member who has been tested and found to harbor a deleterious SNP in a germline tumor-suppressor gene. Testing is indicated for any breast cancer patient with a triple-negative breast cancer, who is <40 years old, who has synchronous or metachronous contralateral breast cancers, who has a personal history of ovarian cancer, or who has a first-degree relative (mother, father, or sister) with breast or ovarian cancer. All males with breast cancer should be tested. Some guidelines suggest testing any breast cancer patient of Ashkenazi descent. Patients with these mutations should be counseled appropriately.

Some experts have recommended testing any patient diagnosed with breast cancer, both for genetic counseling but also because of the advent of effective therapies directed toward cancers that have deleterious *BRCA1/2* mutations (Table 79-1), although this strategy remains controversial. Regardless, any patient who is found to have inherited germline deleterious SNPs should receive formal genetic testing and counseling about special screening and prevention measures they might take.

PREVENTION OF BREAST CANCER

One major reason to determine risk would be to efficiently apply prevention and/or screening strategies, if either has been shown to be effective for the disease of interest. At present, although diet and exercise are certainly recommended approaches to healthy living, none has been proven to specifically decrease a woman’s risk of breast cancer. Avoidance of combined estrogen/progestin HRT reduces the associated increased risk of breast cancer to that of an average woman not using HRT.

Prophylactic removal of the breasts is an effective, albeit drastic, preventive strategy. Bilateral prophylactic mastectomies reduce the risk of breast cancer incidence and mortality by >95%. Because breasts are not encapsulated organs, some normal breast tissue is always left behind, and therefore, women who elect to have prophylactic mastectomies should be counseled that they still have some risk of developing a new breast cancer. Prophylactic mastectomy is most often chosen by women with germline genetic risk, in whom there is evidence of mortality reduction. For women with average or only mildly elevated risks, such as diagnosis of a unilateral breast cancer, survival is not increased by prophylactic mastectomy, and, because of its obvious adverse effect on sexuality, cosmesis, and breast-feeding, this approach is not considered appropriate.

So-called “chemoprevention” to lower breast cancer risk can be achieved with therapies directed toward the ER/estrogen signaling pathway (Table 79-1). These include the selective estrogen receptor modulators (SERMs) as well as aromatase inhibition. The latter should only be applied in postmenopausal women, since aromatase inhibition can result in a paradoxical increase in circulating estrogen levels in women with functioning ovaries. Chemoprevention with SERMs or aromatase inhibition lowers risk of ER-positive breast cancer by approximately one-half, although it has no effect on the more lethal ER-negative breast cancers. Of interest, prophylactic bilateral oophorectomy and salpingo-oophorectomy, which are often performed in women with high genetic risk (such as those with inherited *BRCA1/2* deleterious SNPs), also reduce breast cancer in addition to ovarian cancer risk.

SCREENING FOR BREAST CANCER (FIG. 79-2)

Screening mammography results in earlier diagnosis and subsequent local and systemic therapy. Overviews of nine prospective randomized clinical trials demonstrate that screening mammography reduces breast cancer mortality by one-fifth to one-quarter in women aged ≥ 50 years. The relative reduction in breast cancer mortality for women between ages 40 and 50 years is similar, although the absolute numbers of women who benefit in this age group is smaller since the incidence of breast cancer is much lower in younger women. In addition to reducing breast cancer mortality, screening mammography and early detection are more likely to identify tumors at a stage more appropriate for conservative local therapy. Better technology, including digitized mammography, tomosynthesis, routine use of magnified views, and greater skill in interpretation, have all improved the accuracy of mammography. Magnetic resonance spectroscopy has higher sensitivity but lower specificity than mammography. Since none of these newer technologies has been shown to be superior to mammography in terms of mortality reduction, screening of women with standard risk by any technique other than mammography is not recommended.

The issue of screening breast imaging of any sort has been controversial. Although the prospective randomized clinical trials demonstrate a late reduction in breast cancer mortality, they do not demonstrate

improvement in overall survival. Further, many authors have raised concern about diagnosis of cancers that may be biologically insignificant, raising the specter of overdiagnosis and overtreatment. Moreover, the substantial advances in both local and systemic therapies for breast cancer may have reduced the benefit of earlier diagnosis provided by screening. In contrast, in countries that have adopted widespread screening programs, nonrandomized, epidemiologic studies have demonstrated even greater magnitude reductions in breast cancer mortality than those seen in the randomized trials. Taken together, these data have led most guideline bodies to recommend annual screening for women aged 50–70 years. Many have also recommended screening for women in the 40- to 50-year-old range. For older women, caregiver and patient judgment should be used, taking into account comorbidities.

Magnetic resonance imaging (MRI) is recommended for women with particularly dense breasts, women whose first cancer was not detected by mammography, women with an axillary breast cancer presentation but no definable breast mass on physical exam or mammography, and those with high genetic risk, such as *BRCA1* or *BRCA2* carriers or those with Li-Fraumeni, Cowden's, or Bannayan-Riley-Ruvalcaba syndromes. MRI might also be considered for women with a history of radiation therapy to the chest between ages 10 and 30 years. In these women, the positive predictive value of MRI is higher because

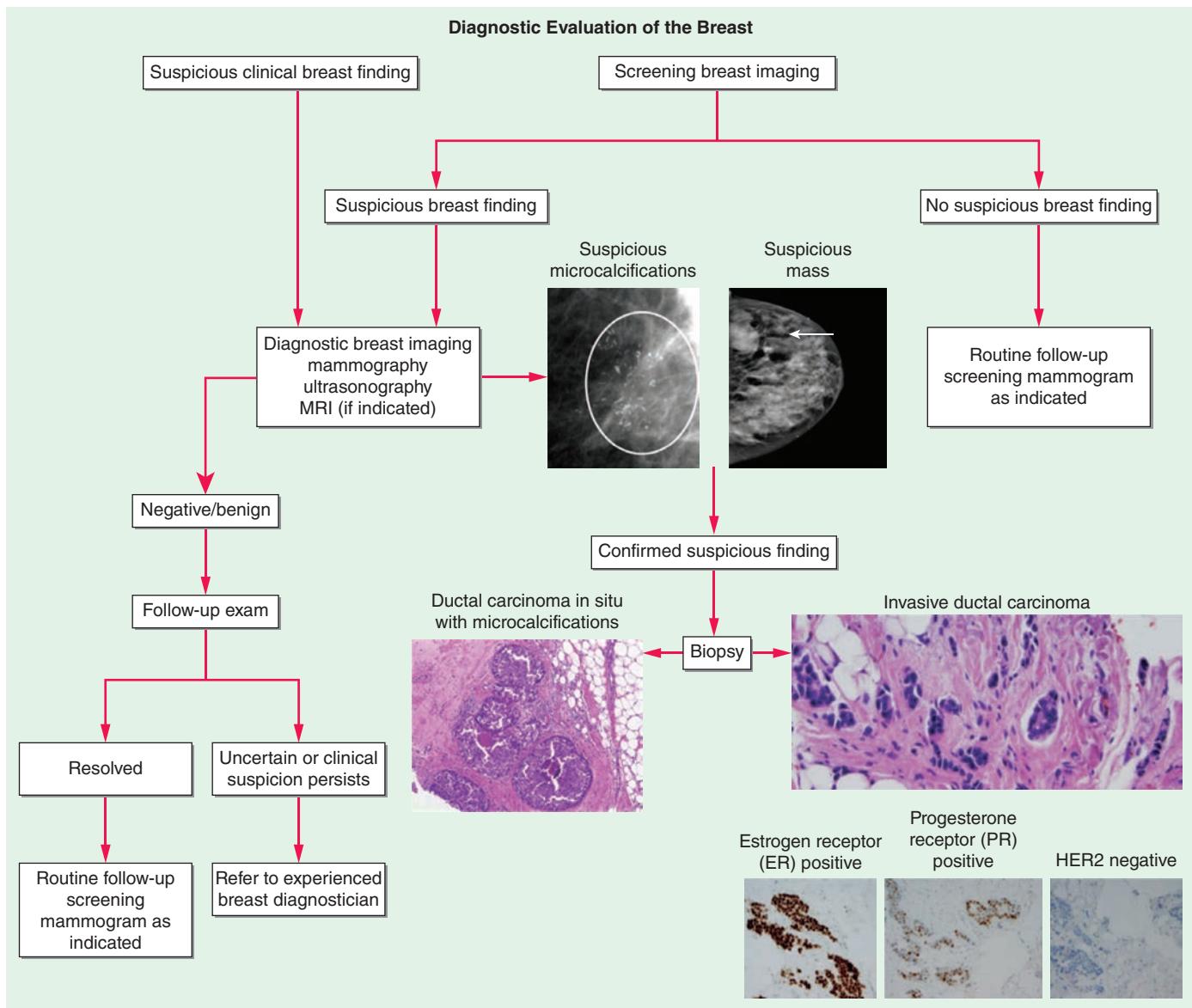


FIGURE 79-2 Evaluation and workup of breast lesions. For more extensive details, see https://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf. (Mammographic images courtesy of Drs. Mark Helvie and Colleen Neal, Department of Radiology, Michigan Medicine. Photomicrographs courtesy of Dr. Celina Kleer, Department of Pathology, Michigan Medicine.)

of the higher incidence of cancer, and furthermore, many of them are considering prophylactic mastectomy as an alternative; therefore, the lower specificity and risk of a false-positive finding has been considered more acceptable.

Self-examination or physical breast examinations done by a health professional have poor sensitivity and specificity, and regular breast self-examination is not recommended. Nonetheless, all women should be familiar with how their breasts normally look and feel and report any changes to a health care provider right away. Because the breasts are a common site of potentially fatal malignancy in women, examination of the breast is an important part of a routine physical examination.

Screening breast imaging is not recommended for men, since it is so unusual and easily detected. It is important to note that unilateral lesions should be evaluated in the same manner as in women with an appropriately high index of suspicion.

EVALUATION OF BREAST MASSES (FIG. 79-2)

Virtually all breast cancer is diagnosed by biopsy of an abnormality detected either on a mammogram or by palpation. The presence or absence of any risk factors, such as age, family history, or menstrual history cannot be used to exclude more careful workup and, if indicated, a biopsy. Any woman with a persistent breast abnormality should be referred to an experienced breast diagnostician in order to avoid delay in diagnosis and therapy.

PALPABLE BREAST MASSES

Proper attention needs to be given to any abnormality either discovered by the patient or appreciated by the health care provider during examination. Most newly diagnosed breast cancers are asymptomatic. Lesions with certain clinical features, including firmness, irregularity, tethering or fixation to the underlying chest wall, and dermal erythema or peau d'orange (skin edema with pockmarking), are very worrisome for breast cancer. In contrast, painful masses and those that are cystic on physical examination are less likely malignant. However, none of these has a high positive or negative predictive value. Likewise, a negative mammogram in the presence of a persistent lump in the breast does not exclude malignancy and, again, deserves careful workup.

In premenopausal women, lesions that are either equivocal or nonsuspicious on physical examination should be reexamined in 2–4 weeks during the follicular phase of the menstrual cycle. Days 5–7 of the cycle are the best time for breast examination. A dominant mass in a postmenopausal woman or a dominant mass that persists through a menstrual cycle in a premenopausal woman requires further evaluation, likely including biopsy if appropriate.

ABNORMAL MAMMOGRAM

Diagnostic mammography, which is performed after a palpable abnormality has been detected, should not be confused with *screening mammography*, which is performed in an asymptomatic woman with no previously identified abnormalities.

Abnormalities that are first detected by physical exam and/or screening mammography should be evaluated by diagnostic mammography. Suspicious mammographic abnormalities include clustered, heterogeneous, linear, and branching microcalcifications; densities (especially if spiculated); and new or enlarging architectural distortion. For some lesions, ultrasound may be helpful either to identify cysts or to guide biopsy. If there is no palpable lesion and detailed mammographic studies are unequivocally benign, the patient should have routine follow-up appropriate to the patient's age. If a nonpalpable mammographic lesion has a low index of suspicion, mammographic follow-up in 3–6 months is reasonable. The presence of a breast lump and a negative mammogram does not rule out cancer, and if the physical finding persists or enlarges during follow-up, further evaluation and, if appropriate, a biopsy are indicated.

BREAST MASSES IN PREGNANCY OR LACTATION

Breast cancer develops in 1 of 3000–4000 pregnancies. The breast grows during pregnancy under the influence of estrogen, progesterone,

prolactin, and human placental lactogen. After delivery and during lactation, breast tissue continues to be under the influence of unopposed prolactin. Therefore, breast examination during these times can be challenging. Nonetheless, development of a dominant mass during pregnancy or lactation should not be attributed to hormonal changes without appropriate diagnostic evaluation. Stage for stage, breast cancer in pregnant patients is no different from premenopausal breast cancer in nonpregnant patients. However, pregnant women often have more advanced disease because the significance of a breast mass was not fully considered and/or because of endogenous hormone stimulation.

PATHOLOGIC FINDINGS OF THE BREAST

BENIGN BREAST HISTOPATHOLOGY

Only ~1 in every 5–10 breast biopsies leads to a diagnosis of cancer, although the rate of positive biopsies varies in different countries and clinical settings due to variable interpretation, medico-legal considerations, and availability of mammograms. The vast majority of benign breast masses are due to fibrocystic changes, a descriptive term for small fluid-filled cysts and modest epithelial cell and fibrous tissue hyperplasia. Women with ductal or lobular cell proliferation (~30% of patients), particularly the small fraction (3%) with atypical hyperplasia, have a fourfold greater risk of developing breast cancer than women who have not had a biopsy, and the risk is even higher if they have an affected first-degree relative. Follow-up breast imaging should be continued, but not on an accelerated or more intense fashion than regularly indicated. Chemoprevention with antiestrogen therapy (SERM or aromatase inhibitor [AI]) should be considered for such patients. Prophylactic mastectomy is not normally indicated. By contrast, patients with a benign biopsy without atypical hyperplasia are at little increased risk and may be followed routinely.

NONINVASIVE BREAST NEOPLASMS

Breast cancer develops as a series of molecular changes in the epithelial cells that lead to ever more malignant behavior (Fig. 79-1). These changes range from malignant cells confined within the basement membrane of the lobule or duct, designated "noninvasive" or more commonly "in situ" carcinoma, to cancer cells that have invaded through the basement membrane into the surrounding normal tissue ("invasive" or "infiltrating" cancer). Increased use of mammography has led to more frequent diagnoses of noninvasive breast neoplasms. These lesions fall into two groups: ductal carcinoma *in situ* (DCIS) (Fig. 79-2) and lobular carcinoma *in situ* (LCIS; or lobular neoplasia *in situ* [LNIS]). The management of both entities is controversial.

Ductal Carcinoma *In Situ* Proliferation of cytologically malignant breast epithelial cells within the ducts is termed *ductal carcinoma in situ* (DCIS). Atypical hyperplasia may be difficult to distinguish from DCIS. In many ways, DCIS is really a "premalignant" condition, but probably at least one-third of patients with untreated DCIS develop invasive breast cancer within 5 years. However, many low-grade DCIS lesions do not appear to progress over many years; therefore, many patients are overtreated. Unfortunately, no reliable methods distinguish patients who require treatment from those who may be safely observed.

Mastectomy is nearly 100% effective in preventing a future breast cancer event in that breast and fundamentally can be considered prophylactic surgery, but is often not required for adequate treatment. No prospective randomized studies have directly compared breast-preserving therapy to mastectomy. However, the nearly 100% 10-year survival rates with the former suggest that it is a satisfactory strategy. Breast-preserving therapy refers to excisional surgery alone with or without breast radiation. However, although survival was identical in the two arms of a randomized trial comparing wide excision plus or minus irradiation, the latter caused a substantial reduction in the local recurrence rate as compared with wide excision alone. Addition of tamoxifen or an AI to any DCIS surgical/radiation therapy regimen further improves local control. However, in the largest trial comparing the two in DCIS, anastrozole did not improve distant disease-free or overall survival compared to tamoxifen.

Several prognostic features may help to identify patients at high risk for local recurrence after either lumpectomy alone or lumpectomy with radiation therapy and, therefore, might provide an indication for mastectomy. These include extensive disease within the breast; age <40; and cytologic features such as necrosis, poor nuclear grade, and comedo subtype with overexpression of HER2. In summary, it is reasonable to recommend breast-preserving surgery for patients who have a localized focus of DCIS with clear margins followed by breast irradiation and tamoxifen or anastrozole. Recently, a multifactorial gene expression assay has been shown to predict risk of recurrence in DCIS treated with breast-preserving surgery alone, but it is not clear that the in-breast risk recurrence rate in patients with low recurrence scores is sufficient to avoid radiation. The decision to irradiate such patients depends on the risk aversion to in-breast recurrence balanced against the risk associated with breast irradiation.

For patients with small, unicentric DCIS, axillary lymph node dissection is unnecessary. However, axillary sentinel lymph node (SLN) evaluation, which is discussed in greater detail below, may be indicated for widespread, larger, or poor grade DCIS or if microscopic invasion is identified on a core biopsy. In such cases, subsequent excision or mastectomy may demonstrate invasive disease on the larger specimen. Since SLN mapping is indicated in such patients, doing so at the time of excision or mastectomy avoids a further surgical procedure at a later date.

Lobular Carcinoma (Neoplasia) In Situ The presence of malignant cells within the lobules is termed *lobular carcinoma* or *neoplasia in situ* (LCIS). LCIS does not usually cause palpable breast masses, nor does it often induce suspicious findings on mammogram. Therefore, it usually is found as an incidental finding during pathologic examination of a breast biopsy performed for some other reason. Unlike DCIS, which is usually confined to a single area in a breast, LCIS is often spread throughout the breast, and it is frequently also found in the contralateral breast.

A diagnosis of LCIS itself does not confer a higher risk of mortality from breast cancer, but it does increase the risk of a subsequent breast cancer. Women with LCIS who do not undergo bilateral prophylactic mastectomy experience a new, invasive cancer in either breast at a rate of approximately 1% per year over at least the next 15–20 years, and probably lifelong. Therefore, LCIS is even more commonly considered a premalignant condition than DCIS, and aggressive local management seems unreasonable. Management options include careful observation with routine mammography and chemoprevention with either a SERM or an AI (for postmenopausal women) for 5 years. Beyond 5 years, such patients should be followed with subsequent annual mammography and semi-annual physical examinations. Bilateral prophylactic mastectomy is an alternative option, although it is no more effective in prolonging survival than the less aggressive approach, and it associated with substantial cosmetic, and perhaps emotional, morbidity.

■ INVASIVE BREAST CANCERS

Invasive breast cancers are of more concern than *in situ* lesions because they harbor the capacity to metastasize and cause substantial morbidity and mortality (Fig. 79-1). Eighty-five percent of invasive breast cancers are ductal in origin (Fig. 79-2), 10% are lobular or mixed ductal/lobular, and the other 5% are made up of so-called “special types” including mucinous or colloid (2.4%), tubular (1.5%), medullary (1.2%), and papillary (1%). Although not universally true, prognosis for the special types tends to be better than standard ductal or lobular cancers.

STAGING AND DIAGNOSTIC CONSIDERATIONS

Cancer staging has been traditionally based on the size of the tumor (T) and the presence or absence of regional nodal (N) and distant metastases (M). More recently, tumor grade and biological characteristics, such as expression of ER and HER2, have been incorporated into staging, making the system quite complex. Staging can be performed clinically or pathologically, before or after adjuvant systemic therapy. These are designated as a prefix before the stage as cTNM or pTNM

if determined before or yTNM if determined after systemic (neoadjuvant) therapy. Although staging is an important part of the surgical evaluation and pathology reporting system, the specific elements that inform the clinician of both prognosis and likelihood of response to specific therapies have become more critical determinants of patient care than a simple stage designation. Importantly, imaging for detection of distant metastases is not needed in a patient with no signs or symptoms of widespread disease and who has a T3 or smaller tumor and fewer than four involved axillary lymph nodes, since the odds of finding distant metastases in such patients are low and the risk of false positives outweighs true-positive findings. Although finding bone marrow micrometastases or circulating tumor cells (cM0(i+)) has been associated with worse prognosis, how to integrate these into routine clinical care has not been determined, and their assessment is not recommended in patients with early-stage disease.

TREATMENT

Early-Stage Breast Cancer

GENERAL CONSIDERATIONS

Goals of Therapy The goal of therapy for breast cancer in patients who do not have obvious evidence of distant metastases (meaning outside the breast, chest wall, and regional lymph nodes) is cure, or at least substantial survival prolongation. For these patients, treatment strategies are divided into primary and systemic considerations. Primary therapies consist of surgical and radiation treatments directed toward the breast and locoregional lymph nodes. These approaches are designed to minimize the odds of locoregional recurrence while maintaining quality of life and cosmesis as much as possible by excising the cancer and sterilizing unaffected breast tissue as appropriate. Adjuvant systemic treatments, consisting of endocrine, anti-HER2, and/or chemotherapies, are given to treat micrometastases that may have already escaped to distant sites but are not yet detectable.

Prognostic and Predictive Factors All treatments for breast cancer are based on prognostic and predictive factors. Prognostic factors provide an indication of how likely a cancer will recur either locally or in distant organs in the future if a patient is not treated with the respective treatments. Predictive factors are used to determine if a given treatment is likely to work or not, assuming the patient's prognosis justifies treatment (or further treatment assuming the patient has been treated in some manner already).

Anatomic prognostic features include visual and physical examination findings of locally advanced breast cancer (T4 lesions: skin erythema [“inflammatory”] or edema [“peau d'orange”], nodules, or ulceration or tumor fixation to the chest wall). In patients without any of these findings, the most important prognostic features remain tumor size (T) and lymph node (N) status.

Biologic features, such as histologic tumor grade as well as ER, PgR, and HER2 status, are also prognostic. Indeed, gene expression patterns, or “signatures,” have demonstrated that breast cancer is actually many different diseases and can be divided into a series of intrinsic subtypes. These subtypes are driven principally, although not exclusively, by expression of ER and HER2 and their respective associated pathways, as well as measures of cellular proliferation and other less important but still contributory biologic features. These intrinsic subtypes are important clinically, both in influencing natural history as well as in prognosis and therapeutic decision making. Four different intrinsic subtypes are recognized: luminal, HER2-like, basal, and claudin-low. Some, if not all, have been further divided into subgroups.

Luminal breast cancers are almost always positive for ER and negative for HER2 amplification. **Luminal A** tumors have the highest levels of ER and downstream related genes, are almost universally negative or low in HER2, are usually low grade, have low proliferative thrust, and have a generally favorable prognosis. They are most likely to respond to endocrine therapy and may appear

to be less responsive to chemotherapy. *Luminal B* breast cancers tend to be PgR negative, may express HER2 but at low levels, are usually higher grade, and have higher proliferative activity than luminal A tumors. Prognosis is somewhat worse than for luminal A cancers, and although not yet proven, they may be more sensitive to chemotherapy.

HER2-amplified breast cancers exhibit co-amplification and overexpression of other genes adjacent to *HER2*. Historically, the clinical prognosis of such tumors was poor, but it has markedly improved with the introduction of targeted anti-*HER2* therapies.

Basal breast cancers are mostly negative for expression of ER/PgR and HER2. Tumors of this type are often called “triple-negative” malignancies, although this is a general term, and such cancers have been further subgrouped based on other genetic abnormalities. They tend to be high grade and express cytokeratins 5/6 and 17 as well as vimentin, p63, CD10, α -smooth muscle actin, and epidermal growth factor receptor (EGFR). Patients with germline *BRCA1* mutations also usually fall within this molecular subtype.

Normal breast-like and *claudin-low* cancers have also been distinguished, but at present, these designations have failed to have clinical significance.

Over the past decade, several multiparameter tests based on gene expression have been developed to determine prognosis in patients who have node-negative, ER-positive, and HER2-negative disease. These assays have been principally used to guide decisions regarding use of adjuvant chemotherapy, as discussed below. Predictive features are usually used to guide targeted systemic therapies. These include ER for endocrine treatments and HER2 for anti-*HER2* therapies, such as trastuzumab, and more recently *BRCA1/2* and *PIK3CA* mutations for poly (ADP ribose) polymerase (PARP) inhibitors and PIK3CA inhibitors, respectively.

LOCAL (PRIMARY) TREATMENTS

In the 1980s, the Halsted radical mastectomy was replaced with the less disfiguring modified radical mastectomy, in which chest wall muscles are preserved and only a sampling of axillary lymph nodes are removed. Subsequently, breast-conserving treatments, consisting of surgical excision of the primary tumor (lumpectomy, quadrantectomy, or partial mastectomy) often followed by locoregional radiation, were introduced and shown to have equal if not slightly superior outcomes to those associated with mastectomy. For women undergoing breast conservation, postlumpectomy radiation is usually indicated, although it may be less necessary in older women with ER-positive, node-negative breast cancer, since their risk of subsequent in-breast recurrence is quite low with surgery and endocrine therapy only. When lumpectomy with negative tumor margins is achieved and radiation is delivered appropriately, breast conservation is associated with a recurrence rate in the breast of $\leq 5\%$.

Not all patients are candidates for breast-conserving therapy. Contraindications include large tumor to breast ratio, inability to achieve clear margins with adequate cosmesis after extensive surgery, multifocal cancers, extensive four-quadrant DCIS, and inability to receive radiation. The latter issue arises in women with dermal autoimmune disease (such as lupus erythematosus), prior radiation to the site, and/or lack of available radiation treatment facilities. Further, although not contraindicated, breast-conserving therapy may be less cosmetically acceptable than mastectomy with reconstruction if the nipple-areolar complex is involved with cancer and must be sacrificed. This is a personal choice, and some women prefer mastectomy, especially those with high genetic risks for second breast cancers.

Enigmatically, in spite of the supporting data, only approximately one-third of women in the United States are managed by lumpectomy. It appears that many women still undergo mastectomy who could safely avoid this procedure and probably would if appropriately counseled. Most patients should consult with an experienced breast surgeon and radiation oncologist before making a final decision concerning local therapy. Indeed, a multimodality clinic in which the surgeon, radiation oncologist, medical oncologist, and

other caregivers cooperate to evaluate the patient and develop a treatment plan is usually considered a major advantage by patients.

For patients who do undergo mastectomy, nipple-areolar-sparing mastectomy preserves the dermis and epidermis of the nipple but removes the major ducts from within the nipple lumen and often provides more acceptable cosmesis when combined with reconstruction. This approach is often a preferable option for patients who are having prophylactic surgery or those with cancer who are candidates for immediate reconstruction. Nipple-sparing mastectomy is contraindicated in the presence of inflammatory breast cancer, clinical involvement of the nipple-areolar complex, nipple retraction, Paget disease, bloody nipple discharge, or multicentricity. The safety of nipple-sparing mastectomy is based on retrospective, nonrandomized cohort series. In a meta-analysis of 20 studies (5594 patients), overall and disease-free survival and locoregional recurrence rates appeared similar to those of patients undergoing modified radical mastectomy.

After mastectomy, breast reconstruction is an acceptable option. Breast reconstruction can be achieved by either placement of an exogenous implant (usually silicon) or by transferring autologous tissue from another site, such as the abdomen, latissimus dorsi, or gluteal areas, to the breast. Of note, patients should be aware that a reconstructed breast is usually insensate. Risks of reconstruction include surgical complications such as infection and hemorrhage. Reconstruction does not hinder detection of future recurrences, nor is silicone implant reconstruction associated with non-cancer-related syndromes, although on occasion, these can rupture and removal is required. Breast implant-associated anaplastic large cell lymphoma is an extraordinarily rare complication of textured silicone implants. Although occasionally associated with metastatic lymphoma, it is usually confined locally and highly curable. The optimal choice of implant reconstruction should be made with an experienced breast plastic surgeon.

Postmastectomy chest wall and regional nodal radiation reduces locoregional recurrence and improves survival. It is indicated for patients with high risk of locoregional recurrence, such as those with tumors ≥ 5 cm, four or more positive axillary lymph nodes, or postoperative positive margins. Postmastectomy radiation is not indicated in women with cancers <2 cm, negative lymph nodes, and negative margins. It is considered for women who fall into the areas between these (2–5 cm, one to three positive nodes, or close margins) and is usually recommended if a patient has one to three involved axillary lymph nodes. Many radiation oncologists and plastic surgeons prefer postmastectomy radiation before reconstruction.

The survival of patients who have recurrence in the breast after proper treatment (adequate surgery and radiation if indicated) is somewhat worse than that of women who do not have in-breast recurrences, but it is better than those who suffer locoregional recurrence after mastectomy. Thus, locoregional recurrence is a negative prognostic variable for long-term survival but not the cause of distant metastasis.

Evaluation and Treatment of the Axillary Lymph Nodes SLN mapping and biopsy (SLNB) is generally the standard of care for women with localized breast cancer and clinically negative axilla. This procedure involves injecting a dye or radioactive tracer into the involved breast and, a few hours (4–24) later, undergoing resection of the axillary node containing the dye or tracer. If that lymph node is negative for tumor, more extensive axillary surgery is not required, avoiding much of the risk of postdissection lymphedema. Even in the presence of sentinel lymph node involvement, further axillary surgery may not be required for selected patients, such as older women and those with ER-positive cancers.

ADJUVANT SYSTEMIC THERAPIES

The use of adjuvant systemic therapy is based on the concept that with increasing generations of cellular replication, genetic abnormalities accumulate. These mutations occur randomly and may lead to sensitivity or resistance to therapies, but of course, the latter

is of greater concern. Almost all patients with metastatic breast cancer are destined to die with, if not of, their cancer. However, treatment with the same therapies administered earlier, in the setting of micrometastatic disease only, is more effective than waiting until symptomatic, documented metastases occur and substantially improves survival. More than half of the women who would otherwise die of metastatic breast cancer remain disease-free and experience considerable survival advantage when treated with the appropriate adjuvant systemic regimen.

Prognostic and Predictive Variables Adjuvant systemic therapies are of three types: (1) chemotherapy; (2) endocrine therapy; and (3) anti-HER2 therapies. The decision of whether to apply adjuvant systemic therapy, and which type, depends on prognostic and predictive features as well as the combined judgment of the patient and caregiver.

Prognostic Factors As noted, prognostic factors help define who most likely needs, or perhaps more importantly does not need, adjuvant systemic therapy. In contrast, predictive factors help identify which therapies are likely to work, independent of prognosis (Table 79-1). The most important prognostic variables are provided by *tumor staging: tumor size (T), lymph node status (N), and detectable distant metastases (M)* (Table 79-2). *Histologic grading* is also important. Tumors with a poor nuclear grade (grade 3) have a higher risk of recurrence than tumors with a good nuclear grade (grade 1). Infiltrating lobular cancer, which is almost always ER positive, has roughly the same prognosis as ER-positive infiltrating ductal cancer, although the lobular subtype may be slightly worse. Lobular cancers are harder to detect on mammography and within axillary lymph nodes than ductal cancers, and when they do metastasize, they often spread to unusual sites, such as mesothelial surfaces, the ovaries, and gastrointestinal organs. Among the special types of breast cancer, pure tubular and mucinous cancers are associated with very favorable prognoses. Medullary cancers are often triple negative with poor nuclear grade, but paradoxically, they have a heavy infiltrating lymphocyte component, and they also have a favorable prognosis. However, before treatment is directed toward these types of cancers, their histology should be confirmed by an experienced breast pathologist.

Adjuvant systemic therapy may not be needed at all for patients with very small (<1 cm) tumors and negative lymph nodes. However, every patient with invasive breast cancer has some risk of subsequent distant metastases. Most patients are more likely to accept endocrine therapy for a very small potential benefit than they would accept chemotherapy for the same calculated advantage because the former is much less often associated with either life-threatening or permanently life-changing toxicities.

The greatest controversy concerns the recommendation for adjuvant *chemotherapy*. Since no established factor predicts sensitivity or resistance for this class of treatments, the decision must be made on prognosis alone. Overall, chemotherapy reduces the risk of recurrence over the 10 years subsequent to primary diagnosis by approximately one-third. For patients with T4 cancers or many positive lymph nodes, the risk of distant recurrence (and thus not

being cured) in the subsequent decade is 50% or higher. Therefore, a one-third reduction of a 50% risk of recurrence means that at least 15–20% ($\text{one-third} \times 50\%$) of women will be cured who would not have been cured in the absence of adjuvant chemotherapy. The life-threatening or permanently life-changing toxicities of adjuvant chemotherapy are ~1–2%, and therefore, almost all medical oncologists would recommend adjuvant chemotherapy in this setting.

In contrast, adjuvant chemotherapy is rarely justified in most women with tumors <1 cm in size whose axillary lymph nodes are negative. However, this decision is very much influenced by the expression of ER and HER2. For example, the risk of recurrence of a patient with a small, node-negative but triple-negative breast cancer over the succeeding 10 years without any adjuvant therapy is 15%. If chemotherapy reduces this risk by approximately one-third or more, then approximately 5% or more of patients will be cured who would otherwise have died of their disease. Likewise, a patient with ER- and PgR-negative but HER2-*positive* disease has a slightly worse prognosis, with a risk of recurrence over 10 years of approximately 20%. She will benefit not only from the adjuvant chemotherapy but also from anti-HER2 therapy, so that her potential absolute benefit is even higher. Many, but not all, clinicians would recommend adjuvant chemotherapy for such patients.

On the other hand, patients with ER-positive disease have a better prognosis than those with ER-negative breast cancer, and adjuvant endocrine therapy will further reduce the odds of recurrence by approximately one-half. Therefore, the same patient in the example above (<1 cm, node negative) but who has an ER-positive and HER2-negative cancer has a lower initial risk of recurrence (~10% over 10 years). She is very likely to accept adjuvant endocrine therapy, further lowering her estimated risk of recurrence to ~5%. Even if chemotherapy reduces this residual risk by approximately one-third, no more than 1–2% ($\text{one-third} \times 5\%$) of patients will benefit. This potential benefit is approximately the same as the number of patients who will suffer life-threatening or permanently life-changing toxicities from chemotherapy. Thus, in this case, most clinicians would recommend adjuvant endocrine therapy but not chemotherapy.

Multiparameter gene expression assays have refined prognostic determination, particularly in node-negative, ER-positive, and HER2-negative breast cancers. These tests include the 21-gene Oncotype DX, the 12-gene Endopredict, the 58-gene ProSigna, and the 2-gene Breast Cancer Index. Furthermore, several investigators have reported that analysis of ER, PgR, HER2, and Ki67 by immunohistochemistry (IHC4) also provides prognostic information in this group, but the analytical validity of this assay is quite variable among different pathologists. Assuming adequate adjuvant endocrine therapy, the prognosis of such patients whose tumors have low recurrence scores, which usually identifies luminal A type cancers, with one of these assays is so good they can safely forego adjuvant chemotherapy. Indeed, the same is true for such patients with intermediate Oncotype DX recurrence scores. In contrast, those with high recurrence scores (>25) appear to have luminal B cancers, and the benefits of adjuvant chemotherapy clearly outweigh the risks.

The largest data set for directing care has been generated using the 21-gene recurrence score. However, only one of these tests should be ordered for a single patient, since they do not always give the same results, and there are no data to determine which, in the case of discordance, might be “correct.” Use of these assays to determine prognosis in patients with higher anatomic stage, such as T3b/T4 lesions, or multiple positive lymph nodes, especially if more than three, is still under investigation.

Several measures of tumor growth rate correlate with early relapse, but their use is problematic due to analytical variability. Of these, assessment using immunohistochemical (IHC) assays for the proliferation marker Ki67 is the most widespread. However, substantial lab-to-lab variability and disagreement regarding optimal cut points exist. At present, in standard practice outside of a highly skilled laboratory, Ki67 expression is not used to make clinical decisions.

TABLE 79-2 5-Year Survival Rate for Breast Cancer by Stage

STAGE	5-YEAR SURVIVAL, %
0	99
I	92
IIA	82
IIB	65
IIIA	47
IIIB	44
IV	14

Source: Modified from data of the National Cancer Institute: Surveillance, Epidemiology, and End Results (SEER).

Predictive Factors The two most important predictive factors in breast cancer are ER and HER2 expression, and they should be performed on all primary or metastatic cancer biopsy specimens (Table 79-1). Adjuvant endocrine therapy reduces the risk of recurrence by one-half or more in patients with ER-rich cancers, whereas no detectable benefit is noted in patients with ER-poor or -negative cancers. ER is expressed as the percentage of positive cells within the cancer after IHC staining. Endocrine therapy is recommended for any patient with $\geq 10\%$ positive cells, but not for those whose cancers only have 0–1% staining. The evidence supporting benefit in cases with 1–9% expression is weak, but given the potential benefit and relatively low toxicities of endocrine therapy, it is recommended for such patients with a low threshold for discontinuation if side effects are intolerable.

The HER2 protein is the target for anti-HER2-directed therapies. Adjuvant trastuzumab therapy reduces the risk of distant recurrence and death in patients with HER2-positive breast cancer by one-third or more but has no discernable effect on HER2-negative cancers. HER2 status is determined using either IHC staining for protein overexpression or fluorescent *in situ* hybridization (FISH) for gene amplification. IHC staining of 3+ (on a scale of 0–3+) is considered positive, whereas 0–1+ is considered negative. For cases with 2+ staining, reflex FISH analysis is recommended. FISH can either be used as the initial evaluation or for additional evaluation in IHC 2+ cases. HER2 is considered amplified if the ratio of HER2 to centromere signal on chromosome 17 is ≥ 2.0 . FISH is unnecessary if IHC is 3+ or 0–1+, nor is there reason for IHC testing if FISH is ≥ 2.0 .

No reliable predictive factors exist for chemotherapy in general or for specific types of chemotherapies. It has been hypothesized that chemotherapy may be more active in ER-negative and/or HER2-positive cancers. Luminal B cancers may be more chemotherapy sensitive, whereas luminal A cancers are perceived to be relatively chemotherapy resistant. At present, none of the tests for intrinsic subtype should be used to determine not to give chemotherapy to patients with poor anatomic *prognosis*, such as those with T4 or multiple positive nodes, based on *prediction* of resistance. Attempts to identify reliable predictive factors for individual classes of chemotherapeutic agents (such as anthracyclines, alkylating agents, or taxanes) have been unsuccessful. The platin salts (carboplatin, cisplatin) may have higher activity in patients with triple-negative breast cancer and perhaps in patients with HER2-positive disease. The PARP inhibitors may be more active in patients whose tumors have defects in homologous recombination DNA repair, a group that includes those with *BRCA* mutations.

Adjuvant Regimens • Endocrine Therapy Adjuvant endocrine therapy is indicated for nearly all patients with a diagnosis of ER-positive breast cancer and never for those with ER-negative disease. Two adjuvant endocrine therapy strategies are proven: the SERM tamoxifen or estrogen ablation. In addition to being effective in preventing new cancers and reducing the risk of locoregional recurrences in patients with DCIS, tamoxifen reduces the risk of distant recurrence and death due to invasive breast cancer by ~40% over the decade following diagnosis. It is equally effective in pre- and postmenopausal women, although it may be slightly less effective in very young (<40 years) patients. Because tamoxifen is a SERM, it has mixed ER antagonism (in the breast and brain) and agonism (in the bone, liver, and uterus). Therefore, it is active against breast cancer in the prevention, adjuvant, and metastatic settings.

Side effects of tamoxifen are predictable based on ER antagonism, including frequent hot flashes as well as vaginal discomfort/sexual dysfunction and myalgias and arthralgias. The agonistic effect results in reduction of osteopenia/osteoporosis, especially in postmenopausal women, but it increases thrombosis risk and endometrial cancers due to this effect in the liver and uterus, respectively.

Estrogen depletion can be achieved surgically in premenopausal women by oophorectomy or ovarian suppression with a gonadotropin-releasing hormone (GnRH) superagonist, such as goserelin or leuproide, which invoke a tachyphylactic response, or a GnRH antagonist, such as triptorelin. However, women with nonfunctioning ovaries, whether induced or by natural menopause, still produce small amounts of estrogen by adrenal synthesis of estrogen precursors (testosterone, dehydroepiandrosterone [DHEA]). These are converted to estradiol and estrone by aromatase activity in peripheral fat and possibly cancer cells. In postmenopausal women, circulating estrogen can be reduced to nearly imperceptible levels with the use of oral AIs: anastrozole, letrozole, and exemestane. The three AIs are not significantly different in activity or toxicity. All are slightly more effective than tamoxifen.

Toxicities of the AIs are predictable based on very low estrogen levels. These include hot flashes, musculoskeletal symptoms, and atrophic vaginitis/sexual dysfunction. They also induce or worsen osteoporosis and fractures, although this effect can be abrogated with bone-modifying agents, such as bisphosphonates or rank ligand antagonists (denosumab).

For both tamoxifen and the AIs, musculoskeletal symptoms mimicking osteoarthritis and arthralgias can be treated with physical therapy and nonsteroidal anti-inflammatory drugs. After a brief period of washout after discontinuation, switching from one AI to another relieves this symptom in approximately a third of patients. These symptoms can also be reduced with either acupuncture or the antidepressant duloxetine. If AIs cannot be tolerated, tamoxifen is a reasonable therapy, assuming no contraindications, such as a past history of thrombosis or high risk of cerebrovascular disease. Hot flashes from either class of drugs are alleviated in approximately one-half of patients with use of one of several different antidepressant drugs.

For premenopausal women, optimal endocrine therapy depends on prognosis and patient choice. Complete estrogen depletion is slightly more effective than tamoxifen alone, but it may also be associated with more bothersome side effects, such as hot flashes, vaginal dryness, and sexual dysfunction. Complete estrogen depletion, consisting of either oophorectomy or chemical suppression of gonadotropins coupled with an AI, is indicated for women with worse prognosis, in particular node positivity. For those with more favorable prognosis, tamoxifen alone or with ovarian suppression is adequate and produces better quality of life. The AIs should not be administered to women with functioning, or dormant, ovaries, since the negative hypothalamic-pituitary feedback can result in a rebound overproduction of ovarian estrogens.

The duration of adjuvant endocrine treatment is unclear. Formerly, the standard recommendation was at least 5 years of therapy, which clearly reduces the risk of recurrence during that time and for a few years after discontinuation. However, the annual risk of distant recurrence during the subsequent 15 years is 0.5–3%, depending on the initial T and N status. Extended adjuvant endocrine therapy with either tamoxifen or an AI for at least 5 more years continues to reduce this late risk of relapse. The decision of whether to continue adjuvant endocrine therapy or not after 5 years must therefore take into consideration initial risk (T, N, grade), current side effects and potential cumulative toxicities, and the patient's perception of the relative and absolute benefits and risks.

Chemotherapy Multiple-agent adjuvant chemotherapy is more effective than single-agent chemotherapy. Although chemotherapeutic agents are usually delivered in combination, sequential single-agent chemotherapy is as effective, and may be slightly less toxic, although it requires longer total duration to deliver. Administration of four to six cycles of chemotherapy appears to be optimal; one cycle is less effective than six, but more than six cycles have generally increased toxicity without further efficacy. Importantly, although chemotherapy is combined with anti-HER2 therapy in patients with HER2-positive cancers, concurrent endocrine therapy, in particular tamoxifen, is antagonistic with chemotherapy.

Therefore, they are administered sequentially, starting the endocrine therapy after completion of chemotherapy.

Several chemotherapeutic agents have activity in the adjuvant setting. These include alkylating agents (principally cyclophosphamide), anthracyclines (doxorubicin, epirubicin), antimetabolites (5-fluorouracil [5-FU], capecitabine, methotrexate), the taxanes (paclitaxel, docetaxel), and the platinum salts (cisplatin, carboplatin). Within classes, randomized trials have failed to demonstrate superiority of one agent versus another (e.g., doxorubicin vs epirubicin, or paclitaxel vs docetaxel). Escalation above an optimal dose is not more effective. The antineoplastic advantage of more frequent scheduling for most individual agents has been demonstrated in a well-done meta-analysis. Weekly or every-other-week paclitaxel is superior to every-3-week infusion, whereas, enigmatically, the opposite is true for docetaxel. Taken together, the data support giving adjuvant chemotherapy in a dose-dense fashion.

The oldest combination regimen consists of cyclophosphamide, methotrexate, and 5-FU (CMF). Addition of an anthracycline or substitution of an anthracycline for the antimetabolites improves outcomes slightly, albeit with slightly increased risk of heart failure and secondary leukemia. Addition of a taxane to an anthracycline-based regimen further modestly reduces the chances of distant recurrence and death. Likewise, addition of an anthracycline to a taxane-based regimen is also modestly more effective than a taxane plus cyclophosphamide alone.

Which regimen is appropriate for a patient must be individualized based on prognosis, comorbid conditions, and the perspective of the patient. For example, the modest relative improvement of giving an anthracycline, cyclophosphamide, and a taxane (AC-T) may not translate to a sufficiently large absolute improvement in survival in a patient with a relatively small (T2) tumor and negative nodes, whereas that same relative reduction in death may translate to a sufficiently large absolute benefit in a patient with a worse prognosis. Therefore, the former patient might best be served with a taxane/cyclophosphamide (TC) regimen alone, while the latter might wish to accept the added risk of congestive heart failure and leukemia associated with the anthracyclines.

Neoadjuvant Chemotherapy Preoperative, or “neoadjuvant,” treatment involves the administration of adjuvant systemic therapy, most commonly chemotherapy, before definitive surgery and radiation therapy. Neoadjuvant endocrine therapy for patients with ER-positive disease is usually given preoperatively for 4–6 months. However, it is generally reserved for patients for whom a reason for surgical delay exists, such as comorbid conditions.

The objective partial and complete response rates of patients with breast cancer to neoadjuvant chemotherapy range from 10 to 75% depending on the intrinsic subtype of the cancer and the regimen used. Thus, many patients will be “downstaged” by neoadjuvant chemotherapy. In this circumstance, patients with locally advanced, inoperable cancers may become candidates for surgery, and approximately 15% of patients who are not considered eligible for breast-conserving surgery may become so due to shrinkage of their cancer. However, overall survival has not been improved using this approach as compared with the same drugs given postoperatively.

Patients who achieve a pathologic complete remission (pCR) after neoadjuvant chemotherapy have a substantially improved survival compared to those who do not. It is unknown if this observation implies that the latter group did not benefit or just had a worse initial prognosis, yet still gained some benefit. Delivering more therapy to patients who do not have a pCR is appealing. However, it is possible that these patients have chemotherapy-resistant disease, and therefore, more chemotherapy may not be of value. Clearly nonchemotherapeutic strategies, such as adjuvant endocrine therapy if they have an ER-positive breast cancer and adjuvant anti-HER2 therapy if their cancer is HER2 positive, are warranted.

Adding or changing systemic therapies may benefit selected groups of patients who do not have a pCR. Approximately 6 months

of a postsurgical oral fluoropyrimidine, capecitabine, reduces distant metastases in patients with triple-negative breast cancer who have residual disease after non-fluoropyrimidine-containing neoadjuvant chemotherapy. Similarly, postoperative therapy with an antibody-drug conjugate consisting of trastuzumab and the antitubulin emtansine (ado-trastuzumab emtansine) is superior to continuing unconjugated trastuzumab in patients with HER2-positive breast cancer who did not achieve pCR with preoperative chemotherapy and trastuzumab.

Chemotherapy Toxicities Chemotherapy is associated with nausea, vomiting, and alopecia in nearly 100% of patients. Nausea and vomiting are usually well controlled with modern antiemetics. Small but convincing studies have suggested that the strategy of constricting blood flow to the scalp with various means of cooling is commonly effective in sparing hair loss, without evidence of increased scalp metastases.

More importantly, chemotherapy causes potential life-threatening or life-changing toxicities in 2–3% of all treated patients. These include neutropenia and fever, with a risk of infection of ~1%, which can be prevented with appropriate use of the growth factor filgrastim. Secondary myelodysplasia and leukemia occur in ~0.5–1% of patients treated with anthracyclines as well as with high cumulative doses of cyclophosphamide, usually occurring within 2–5 years of treatment. The anthracyclines cause cumulative dose-related congestive heart failure, which occurs in ~1% of patients treated with standard four to five cycles at 60 mg/m². Peripheral neuropathy is the major dose-limiting and life-changing toxicity of the taxanes. Neuropathy occurs during treatment in ~15–20% of patients, and permanent, chronic neuropathy persists in 3–5%. Many patients complain of cognitive dysfunction, so-called “chemo-brain.” Although occasional cases of apparent organic chemotherapeutic toxic effects on cognitive function are noted, much of this syndrome may be due to anxiety, depression, and fatigue caused by the diagnosis itself or the treatment for it. Although not always, cognitive functioning usually returns to age-adjusted baseline several months after discontinuation of therapy.

Anti-HER2 Therapy The humanized anti-HER2 monoclonal antibody trastuzumab decreases both risk of recurrence and mortality in early-stage breast cancer. Trastuzumab is optimally delivered concurrently with chemotherapy, particularly in association with a taxane. Concurrent treatment with an anthracycline is generally avoided, since the main toxicity of trastuzumab is cardiac dysfunction, which appears more often when the agent is delivered simultaneously with doxorubicin. In patients with reasonably favorable prognosis (T1 or T2, node negative), single-agent paclitaxel plus trastuzumab is an adequate regimen. The addition of a second anti-HER2 monoclonal antibody, pertuzumab, in combination with trastuzumab is modestly superior to trastuzumab alone. When given in the neoadjuvant setting, this combination results in higher pCR rates than single-agent trastuzumab. At least in patients with poor prognostic features, such as positive axillary lymph nodes, the combination significantly reduces distant metastases and perhaps mortality. As noted, neoadjuvant studies have demonstrated that postoperative ado-trastuzumab emtansine is superior to trastuzumab in patients who do not achieve a pCR.

Trastuzumab is administered intravenously weekly or every 3 weeks. Twelve months of trastuzumab therapy are optimal with no additional benefit beyond 12 months. Treatment for 6 months is more effective than no trastuzumab therapy but is inferior to 12 months. A preparation of trastuzumab for subcutaneous injection has been approved by the U.S. Food and Drug Administration.

Selected anti-HER2 tyrosine kinase inhibitors have activity against HER2-positive breast cancer, but their benefit in the adjuvant setting is limited. Lapatinib does not add to trastuzumab therapy, and single-agent adjuvant lapatinib is inferior to single-agent trastuzumab. Another anti-HER2 tyrosine kinase inhibitor, neratinib, is modestly superior to no anti-HER2 therapy. Neratinib has

not been compared to trastuzumab either as a single agent or in combination.

Toxicities of Anti-HER2 Adjuvant Therapies In general, the anti-HER2 therapies are safe and effective. Occasionally patients experience allergic reactions to an initial cycle of trastuzumab, but these usually do not recur. Trastuzumab can cause cardiac muscle dysfunction, although it is rare to observe symptomatic congestive heart failure from adjuvant trastuzumab. Baseline and serial echocardiographic monitoring is indicated. Patients with a past history of cardiac abnormalities should not receive trastuzumab or should be followed and treated by a cardiologist with experience in this condition. Pertuzumab is associated with loose stools and diarrhea, which can usually be managed with antidiarrheal therapy, such as loperamide. The chemotherapy payload of ado-trastuzumab emtansine can cause thrombocytopenia and peripheral neuropathy.

Skeletal Strengthening Agents Bone-strengthening agents that are commonly used to treat osteoporosis, specifically the bisphosphonates, have some limited activity in preventing recurrent breast cancer to bone, particularly in postmenopausal women. In addition, bisphosphonate therapy also reduced breast cancer mortality in this subgroup. The benefit is not significantly associated with any specific bisphosphonate class, treatment schedule, ER status, nodal status, tumor grade, or concomitant chemotherapy. As expected, bone fractures are reduced (relative risk [RR] 0.85; 95% confidence interval [CI] 0.75–0.97; $2 p = .02$). Joint guidelines from the American Society of Clinical Oncology and Cancer Care Ontario recommend “that, if available, zoledronic acid (4 mg intravenously every 6 months) or clodronate (1,600 mg/d orally) be considered as adjuvant therapy for postmenopausal patients with breast cancer who are deemed candidates for adjuvant systemic therapy. Further research comparing different bone-modifying agents, doses, dosing intervals, and durations is required.” The rank-ligand inhibitor denosumab does not prevent relapse in bone or other sites, nor does it reduce mortality.

Novel Adjuvant Systemic Agents Other exciting adjuvant strategies are being tested (Table 79-1). These include PARP inhibitors (olaparib, talazoparib) in patients with known germline *BRCA1* or *BRCA2* mutations or those with triple-negative cancers that share similar defects in DNA repair in their etiology. Likewise, the mTOR inhibitor everolimus and the CDK4/6 inhibitors (palbociclib, ribociclib, abemaciclib) are being tested in the adjuvant setting in combination with antiestrogen therapy. The remarkable results of immune checkpoint inhibitors in other cancers have led to studies of this approach in both metastatic and post-neoadjuvant chemotherapy settings. Their activity with chemotherapy in triple-negative disease appears promising.

STAGE III BREAST CANCER

Ten to 25% of patients present with so-called locally advanced or stage III breast cancer at diagnosis. Many of these cancers are technically operable (T3), whereas others, particularly cancers with chest wall involvement, inflammatory breast cancers, or cancers with large matted axillary lymph nodes (T4 or N2-3), cannot be managed with surgery initially. Neoadjuvant downstaging facilitates local therapy. Radiotherapy either to the chest wall after mastectomy or to the breast after tumor excision is almost always recommended, as is regional lymph node treatment. Adjuvant anti-HER2 and endocrine therapies are also used as appropriate. These patients should be managed in multimodality clinics to coordinate local and systemic therapies. Such approaches produce long-term disease-free survival in ~30–50% of patients.

SIMULTANEOUS NEW PRIMARY WITH DETECTABLE METASTASES

In the screening era, only a small fraction of patients (~5%) present with a new primary lesion and simultaneous metastases, detected either due to symptoms of distant disease or because they had staging scans due to locally advanced disease. Several retrospective single-institutional experiences have suggested that neoadjuvant systemic therapy followed by local therapy (breast surgery and

radiation) is associated with prolonged survival. However, two prospective randomized trials have failed to demonstrate any survival benefit. Currently, local therapy for such patients is considered on a case-by-case basis depending on the response to systemic therapy and the patient's overall performance status and desires.

BREAST CANCER SURVIVORSHIP ISSUES

The odds of surviving breast cancer have increased dramatically over the past 35 years due to a combination of early detection and more effective therapies. Without these advances, >60,000 American women would have suffered breast cancer mortality in 2020, and over one-quarter million women are alive who would not have been otherwise. Coupled with the women who would have been cured even before the impressive advances of the past three decades, millions are breast cancer survivors. Thus, all clinicians, not just oncologists, need to be aware of survivorship issues in patients with previously diagnosed and treated breast cancer.

At present, no special follow-up procedures, such as serial circulating tumor biomarkers or systemic radiographic/scintigraphic imaging, are indicated in an asymptomatic patient with no physical findings of recurrence. Although randomized trials have demonstrated slightly higher incidence of detection of metastases with lead times of 3–12 months by surveillance of asymptomatic patients compared to no special follow-up, no evidence suggests that earlier detection improves overall survival. If anything, such surveillance may worsen quality of life due to higher anxiety levels associated with the testing and toxicities associated with earlier treatment in patients who were otherwise doing well at that time.

However, the risk of late metastases in breast cancer survivors is small but real, especially those who had ER-positive disease. These patients remain at risk for distant recurrence for at least 20 years after initial diagnosis, and probably lifelong. The annual risk is relentless, ranging from 0.5% per year for patients with initially negative lymph nodes and grade 1 tumors <1 cm to as high as 3% per year for those who initially had multiple positive lymph nodes. Therefore, especially in patients with prior ER-positive cancers, the physician must carefully assess and evaluate new symptoms, considering whether they might be due to the cancer, the treatment, or an unassociated condition. Judgment needs to be used to decide if blood tests or imaging are required in order to avoid missing a lesion for which appropriate treatment would improve the patient's quality of life but also to diminish overtreatment, with associated inconvenience, anxieties, false-positive results, and cost. Serial echocardiography should be performed every 3 months for patients on adjuvant trastuzumab, but not after it is discontinued.

Several observations suggest that perhaps the recommendations not to do intensive surveillance in patients without signs or symptoms of recurrence might need to be reconsidered. First, the unrelenting annual incidence of long-term distant recurrence for patients with ER-positive disease demonstrates that none of these patients can ever be considered free of risk of metastases. Second, available diagnostic tests have become substantially more sophisticated in the past decade. These include the advent of liquid biopsies beyond just circulating protein markers, such as circulating tumor DNA and circulating tumor cells, as well as more sensitive and specific scintigraphic and imaging techniques, such as positron emission tomography. Finally, the identification of several highly effective targeted therapies, including new endocrine, anti-HER2, and other therapies, provides opportunities to deliver more beneficial and less toxic therapies than the few chemotherapeutic and endocrine agents that were available at the time the older randomized trials were performed (Table 79-1). Ongoing trials are addressing whether incorporating these new technologies and treatments might improve survival as opposed to waiting for emerging symptoms to initiate additional treatment strategies. At present, no clear answers are apparent.

Likewise, serial monitoring for long-term, life-threatening toxicities associated with chemotherapy, such as myelodysplastic syndromes or congestive heart failure, is not warranted since these are quite uncommon and likely to cause obvious symptoms requiring proper evaluation if they occur.

For patients on endocrine therapy, quality-of-life issues may be critical, including hot flashes, sexual difficulties, musculoskeletal complaints, and risk of osteoporosis. Although estrogen therapy, given orally, transdermally, or transvaginally, effectively reduces these side effects, careful consideration should be given for estrogen replacement therapy to these patients because it may counteract the efficacy of the endocrine therapy. Locally administered therapies are often very effective and likely less risky. Nonhormonal treatments, such as selected antidepressants for hot flashes and musculoskeletal symptoms, and counseling and water-based lubricants for sexual issues can be quite helpful. It is important to screen bone density in patients on an AI more frequently than is recommended for the average postmenopausal woman, since total estrogen depletion results in enhanced risk of osteoporosis and risk of fracture. All women should be counseled to take daily calcium and vitamin D replacement, and if osteoporosis is present or osteopenia is worsening, bone-strengthening agents should be administered.

METASTATIC DISEASE

Diagnostic Considerations (Fig. 79-3) About 15–20% of patients treated for localized breast cancer develop metastatic disease in the subsequent decade after diagnosis. Soft tissue, bone, and visceral (lung and liver) metastases each account for approximately one-third of sites of initial relapses. However, by the time of death, most patients will have bone involvement. Recurrences can appear at any time after primary therapy, but at least half occur >5 years after initial therapy, especially in patients with ER-positive disease. A variety of host factors can influence recurrence rates, including depression and central obesity, and these diseases should be managed as aggressively as possible.

For patients with no prior history of metastases, a biopsy of suspicious physical or radiographic lesions should be performed for confirmation that the lesion does represent recurrent cancer. One should not assume that an apparent abnormality is a breast cancer metastasis. Many benign conditions, such as tuberculosis, gallstones, sarcoidosis,

hyperparathyroidism, or other nonmalignant diseases, can mimic a recurrent breast cancer and are of course treated much differently. Moreover, if biopsy is positive for metastases, re-evaluation of ER and HER2 is indicated, since these can differ between the primary and metastatic lesions in up to 15% of cases. Analysis for PIK3CA mutations should be performed if the cancer is ER positive. Predictors of immune checkpoint inhibitor susceptibility, such as PD-L1 expression, should be determined in triple-negative metastatic breast cancers (Table 79-1). Many experts are also recommending some form of next-generation sequencing of all metastatic cancers from any site, although this recommendation is controversial.

Once a recurrence/metastasis is established, some form of body imaging should be performed—either a scintigraphic bone scan and chest and abdomen CT scan or a positron emission tomography (PET)/CT scan, depending on caregiver's preference. Brain scanning (CT or MRI) is not indicated in the absence of any cognitive or neurologic signs or symptoms in most patients. However, because of increased risk of brain metastases in HER2-positive breast cancer, some experts do recommend central nervous system (CNS) imaging in such patients even in the absence of clinical indications. Regardless, body scans provide a perspective of extent of disease, which may guide therapeutic decisions, as well as the need for ancillary treatments, such as bone-modifying agents if skeletal metastases are present.

Considerations Regarding Goals of Therapy Although treatable, metastatic disease is rarely if ever cured. The median survival for all patients diagnosed with metastatic breast cancer is <3 years, but with remarkable variability depending on intrinsic subtype and treatment effects. Patients with triple-negative metastatic breast cancer have the shortest expected survival, whereas those with ER-positive disease can expect to live the longest. HER2 positivity was initially found to be a very poor prognostic factor in metastatic breast cancer, but the availability of several effective targeted treatments has improved the expected survival rates to at least those of ER-positive patients, if not better.

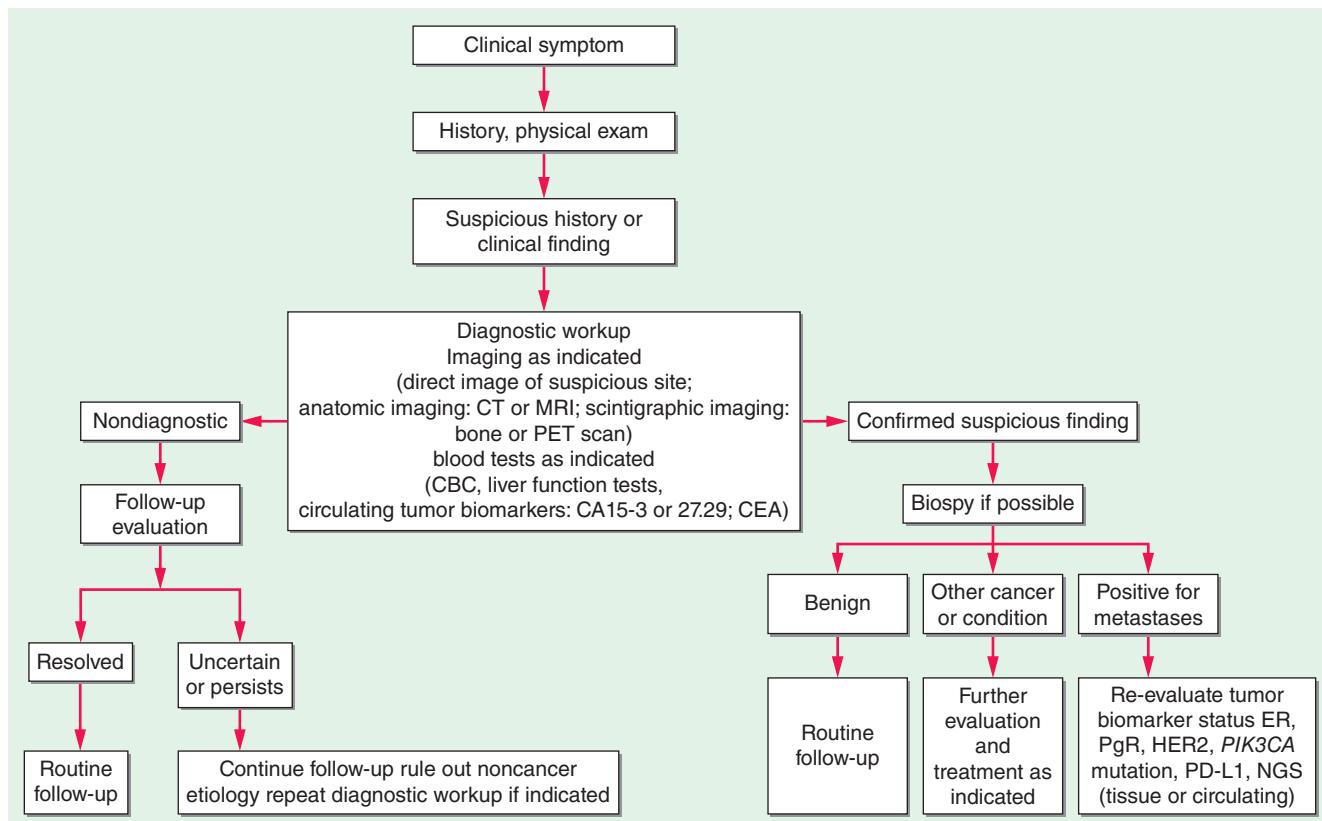


FIGURE 79-3 Evaluation of new signs or symptoms in a patient with prior history of early-stage breast cancer. See text for details. CBC, complete blood count; CEA, carcinoembryonic antigen; ER, estrogen receptor; NGS, next-generation sequencing; PET, positron emission tomography; PgR, progesterone receptor.

The overall goal of treatment of metastatic disease is palliation or, put simply, to “keep the patient feeling as well as she can for as long as she can.” A secondary goal is improved survival. Overall survival has not been improved by advocating more aggressive or toxic therapies, such as high-dose or combination chemotherapy, but rather by using more selective and biologically based therapy, including endocrine or anti-HER2 therapies in patients with ER- or HER2-positive breast cancers, respectively.

Generally, a new treatment is continued until either progression or unacceptable toxicities are evident. These are both evaluated by serial history and physical examinations and periodic serologic evaluation for hematologic or hepatic abnormalities, as well as circulating tumor biomarker tests (assays for MUC1 [CA15-3 or CA27.29] and for carcinoembryonic antigen [CEA] and occasionally CA125). If all these evaluations fail to suggest progression, it is unlikely that imaging will contribute. However, if one or more of these suggest progression, whole-body imaging with whichever modality(ies) was used at baseline is indicated.

The choice of therapy requires consideration of local therapy needs, specifically surgical approaches to particularly worrisome long-bone lytic lesions or isolated CNS metastases. New back pain in patients with breast cancer should be explored aggressively on an emergent basis, usually with a spine MRI; to wait for neurologic symptoms is a potentially catastrophic error. Metastatic involvement of endocrine organs can occasionally cause profound dysfunction, including adrenal insufficiency and hypopituitarism. Similarly, obstruction of the biliary tree or other impaired organ function may be better managed with a local therapy than with a systemic approach. Radiation as an adjunct to or instead of surgery is an important consideration for particularly symptomatic disease in long or vertebral bones, locoregional recurrences, and CNS metastases. In many cases, systemic therapy can be withheld while the patient is managed with appropriate local therapy.

Aggressive local treatment, such as excision, radiation, radiofrequency ablation, or cryotherapy of metastases to the lung, liver, or other distant sites, does not improve survival. Although appealing, these strategies are associated with increased toxicity and cost and should be reserved for palliation.

Locoregional recurrence on the chest wall or surrounding lymph nodes is an exception to this principle. Some of these lesions may well represent new primary cancers, even in the case of prior mastectomy, since some residual at-risk normal tissue can remain. Regardless, rendering the patient disease-free by surgery and radiation, if appropriate, followed by adjuvant systemic therapy, such as endocrine therapy if the cancer is ER positive, or chemotherapy if ER negative, is indicated. Anti-HER therapy is also appropriate if the cancer is HER2 positive.

Selection of the systemic therapy strategy depends on the overall medical condition of the patient, the hormone receptor and HER2 status of the tumor, and clinical judgment. Because therapy of systemic disease is palliative, the potential toxicities of therapies should be balanced against expected response rates. Several variables influence the response to systemic therapy. For example, the presence of ER and PgR is a strong indication for endocrine therapy, even for patients with limited visceral (lung/liver) disease. On the other hand, patients with short disease-free intervals or rapidly progressive visceral disease (liver and lung) with end-organ dysfunction, such as lymphangitic pulmonary disease, are unlikely to respond to endocrine therapy.

Many patients with bone-only or bone-dominant disease have a relatively indolent course. Because the goal of therapy is to maintain well-being for as long as possible, emphasis should be placed on avoiding the most hazardous complications of metastatic disease, including pathologic fracture of the axial skeleton and spinal cord compression. Under such circumstances, systemic chemotherapy has a modest effect, whereas radiation therapy may be effective for long periods. Patients with bone involvement should receive concurrent bone-strengthening agents, such as bisphosphonates or the humanized monoclonal anti-RANK ligand antibody denosumab.

These therapies have been proven to reduce bone pain, fractures, and hypercalcemia of malignancy.

Many patients are inappropriately treated with toxic regimens into their last days of life. Often, oncologists are unwilling to have the difficult conversations that are required with patients nearing the end of life, and not uncommonly, patients and families can pressure physicians into treatments with very little survival value. Although systemic therapy is designed to deliver palliation, formal palliative care consultation and realistic assessment of treatment expectations need to be reviewed with patients and families. We urge consideration of formal palliative care consultations for patients who have received at least two lines of therapy for metastatic disease.

SYSTEMIC TREATMENTS FOR METASTATIC BREAST CANCER

Endocrine Therapy (Table 79-1) Approximately 30–70% of patients with ER-positive breast cancer will benefit from endocrine therapy. Potential endocrine therapies are summarized in Table 79-1. Available strategies include SERMs (tamoxifen, toremifene), the AIs (anastrozole, letrozole, exemestane), and the selective estrogen receptor downregulator (SERD) fulvestrant. Additive endocrine therapies, including treatment with progestins and androgens and, enigmatically, pharmacologic doses of estrogens, are all active, but they may be associated with unacceptable side effects in many women and are rarely used. Tamoxifen withdrawal (as well as withdrawal of pharmacologic doses of estrogens) induces responses in ~15% of patients, but with the advent of so many other therapies for metastatic disease, this strategy is also rarely used in modern oncology.

The sequence of endocrine therapy is variable. Patients who respond to one endocrine therapy have at least a 50% chance of responding to a second endocrine therapy. It is not uncommon for patients to respond to two or three sequential endocrine therapies. Many, but not all, women with ER-positive breast cancer who suffer a recurrence will do so either while still taking or after recently discontinuing a prior adjuvant endocrine therapy (either tamoxifen or an AI). In most postmenopausal patients, if they have never received an AI or discontinued adjuvant AI many years before recurrence, the initial endocrine therapy should be an AI rather than tamoxifen. As noted, AIs are not used in women with functioning ovaries because their hypothalamus can respond to estrogen deprivation by producing gonadotropins that promote ovarian estrogen synthesis. Fulvestrant is usually used in sequence after AI therapy. Compared to single-agent therapy, combination endocrine therapies increase the chances of response, but they do not appear to increase the ultimate time to chemotherapy use or overall survival. Combinations of chemotherapy with endocrine therapy are not useful.

Over the past decade, several different targeted agents have been shown to enhance outcomes of patients with ER-positive metastatic breast cancer when combined with endocrine therapy (Table 79-1). Addition of an inhibitor of mTOR, everolimus, to endocrine therapy improves time to progression. Everolimus is commonly associated with mucositis, which can be prevented or alleviated by use of dexamethasone-containing mouthwash. Diarrhea is also a common side effect and can be lessened with antidiarrheal medications such as loperamide.

Inhibitors of CDK4/6 (palbociclib, ribociclib, abemaciclib) also substantially improve progression-free survival, and even overall survival when combined either with an AI or fulvestrant (Table 79-1). Most experts now recommend a CDK4/6 inhibitor with endocrine therapy as first-line therapy for ER-positive metastatic disease. They can cause dangerous neutropenia, although rarely to the extent seen with chemotherapy. Nonetheless, absolute neutrophil counts need to be monitored closely with appropriate adjustments in dose and schedule. Fatigue is also an occasional side effect, and abemaciclib frequently causes diarrhea. Similarly, an inhibitor of PIK3CA protein, alpelisib, prolongs progression-free survival in patients whose cancers harbor activating mutations of this gene. Like everolimus, it too causes mouth sores and diarrhea.

These targeted agents should not be given simultaneously but rather in sequence as appropriate.

Chemotherapy Unlike many other epithelial malignancies, breast cancer responds to multiple chemotherapeutic agents, including anthracyclines, alkylating agents, taxanes, and antimetabolites. Multiple combinations of these agents have been found to improve response rates somewhat, but they have had little effect on duration of response or survival. Unless patients have rapidly progressive visceral (lung, liver) metastases with end-organ dysfunction, single-agent chemotherapy is preferable, used in sequence as one drug fails going on to the next. Given the significant toxicity of most drugs, the use of a single-agent therapy will minimize toxicity by sparing the patient exposure to drugs that would be of little value. No method to select the drugs most efficacious for a given patient has been demonstrated to be useful.

Most oncologists use capecitabine, an anthracycline, or a taxane for first-line chemotherapy, either in a patient with ER-positive disease that is refractory to endocrine therapy or for a patient with ER-negative breast cancer. Within these general classes, one particular agent is no more preferable than another (such as doxorubicin vs epirubicin or paclitaxel vs docetaxel), and the choice has to be balanced with individual needs. Objective responses in previously treated patients may also be seen with gemcitabine, vinorelbine, and oral etoposide, as well as a newer class of agents, ephotilones. Platinum-based agents have become far more widely used in both the adjuvant and advanced disease settings for some breast cancers, particularly those of the triple-negative subtype.

Anti-HER2 Therapy (Table 79-1) Initial use of a trastuzumab, either alone or with chemotherapy, improves response rate, progression-free survival, and even overall survival for women with HER2-positive disease. Indeed, anecdotal reports suggest that, on occasion, a few patients with HER2-positive metastatic breast cancer may be cured. Addition of pertuzumab to trastuzumab is more effective than trastuzumab alone. The antibody-drug conjugate, ado-trastuzumab emtansine, is effective after progression on trastuzumab. Another antibody-drug conjugate, fam-trastuzumab-deruxtecan-nxki, has been shown to be active even in patients who have progressed on multiple other anti-HER2 therapies, including ado-trastuzumab emtansine. A monoclonal antibody, margetuximab, has been engineered to specifically enhance antibody-dependent cell-mediated cytotoxicity against tumor cells overexpressing HER2. In a phase 3 trial, margetuximab plus chemotherapy improved overall survival by 1.8 months compared with trastuzumab plus chemotherapy.

Inhibitors of the HER2 tyrosine kinase domain also have activity against HER2-positive breast cancers. Lapatinib is effective when added to chemotherapy after patients progressed on trastuzumab. In addition, even after progression on trastuzumab, combination trastuzumab and lapatinib is superior to lapatinib alone. When added to oral capecitabine, neratinib is more effective than lapatinib in patients who received two or more prior anti-HER2-based regimens. In patients with heavily pretreated HER2-positive disease, including those with brain metastases, adding tucatinib to trastuzumab and capecitabine resulted in better progression-free survival and overall survival outcomes than adding placebo. Of interest, 2–3% of breast cancers that do not amplify or overexpress HER2 contain activating mutations in the gene encoding for it. Preclinical models and preliminary trials suggest that neratinib is particularly active against this mutation.

PARP Inhibitors (Table 79-1) PARP inhibitors induce synthetic lethality of cancer cells with inactive *BRCA1/2* or cancers that have *BRCA*-like biology by virtue of an ineffective homologous recombination DNA repair mechanism. Both olaparib and taloparib have been approved for patients whose cancers have developed in the context of germline *BRCA1/2* mutations. Both agents, given as a single agent, are as effective as standard chemotherapy, but in general, they are less toxic. Unfortunately, responses are relatively short lived. PARP inhibitors are now being investigated in combination with

chemotherapy and with immune checkpoint inhibitors. PARP inhibitors can cause mild nausea and occasional vomiting as well as fatigue.

Sacituzumab Govitecan (Table 79-1) An antibody-drug conjugate, sacituzumab govitecan, showed activity in a nonrandomized trial of patients with triple-negative metastatic breast cancer. Sacituzumab govitecan combines a humanized immunoglobulin G antibody targeted against TROP-2SN-38 with the active metabolite of irinotecan. In a single-arm, phase 2 trial, this agent elicited responses in one-third of such patients.

Immune Checkpoint Inhibitors (Table 79-1) These therapies permit immune effector cells to recognize and eliminate host cancer cells based on their recognition of neoantigen expression in tumor cells due to chromosomal instability and accumulated mutations. The excitement over immune checkpoint inhibitors has spread to metastatic breast cancer, especially of the triple-negative subtype. Atezolizumab in combination with nab-paclitaxel improves progression-free and perhaps overall survival, although exclusively against cancers with infiltrating immune cells that express PD-L1 (Table 79-1). The side effects of these agents can be life threatening, consisting of induction of inflammatory autoimmune responses in nearly every organ imaginable. These include thyroiditis, pneumonitis, myo- and pericarditis, esophagitis, gastritis and colitis, hepatitis, pancreatitis, hypophysitis, and dermatitis. The endocrine toxicities tend to be irreversible. Careful management should be handled by an experienced team.

Bone-Modifying Agents Bone-modifying agents, such as bisphosphonates or the anti-RANK antibody denosumab, are recommended for all patients with bone metastases. These agents substantially reduce the incidence of cancer-related skeletal events, such as bone pain, fracture, and hypercalcemia of malignancy. The bisphosphonates may cause myalgias and skeletal pain lasting a few hours to days after infusion. Both strategies have been associated with osteonecrosis of the jaw. The incidence of this complication is reduced by ensuring adequate dentition before treatment and by delivering treatment every 3 months, instead of monthly. The former has been shown to have equal efficacy.

BREAST CANCER IN PREGNANCY

As noted, breast cancer is unusual during pregnancy but does occur. Because of pregnancy-related physical breast changes, diagnosis is frequently delayed. Workup is the same as for non-pregnancy-related breast cancers, except radiographic staging should be limited or avoided, especially of the abdomen. Prognosis is similar, stage for stage, as that for age-matched women who are not pregnant. Pregnancy termination is usually not required. However, it is strongly advised that the patient be referred to a high-risk pregnancy program.

Remarkably, adjuvant chemotherapy including doxorubicin and cyclophosphamide can be safely given beyond the first trimester. Taxanes and the platinum salts may be administered safely. In contrast, anti-HER2 antibody therapies have resulted in unacceptable fetal malformations and pregnancy complications and should be avoided. Likewise, endocrine therapies should be delayed until after delivery. In general, a reasonable strategy is to deliver combination neoadjuvant chemotherapy with either concurrent or sequential single agents to permit sufficient embryogenesis followed by delivery and then breast primary therapy (surgery/radiation). Further adjuvant therapies, including additional chemotherapy and/or anti-HER2 and endocrine therapies can be delivered postoperatively. Breast feeding is discouraged, since these agents may cross into milk.

MALE BREAST CANCER

Breast cancer is ~1/150th as frequent in men as in women; ~2000 men developed breast cancer annually in the United States. Risk factors include inherited deleterious SNPs in *BRCA2*, as well as increased exposure to endogenous or exogenous estrogen. Men with Klinefelter's syndrome have two or more copies of the X chromosome and have higher levels of estrogen. Other conditions of hyperestrogenism, such as in hepatic failure and with exogenous

estrogen use in transgender situations, are also associated with higher risk of male breast cancers. However, the vast majority of men who present with breast cancer have none of these conditions.

Breast cancer usually presents in men as a unilateral lump in the breast and is frequently not diagnosed promptly. Given the small amount of soft tissue and the unexpected nature of the problem, locally advanced presentations are somewhat more common. Although gynecomastia may initially be unilateral or asymmetric, any unilateral mass in a man >40 years old should be biopsied. On the other hand, bilateral symmetric breast development rarely represents breast cancer and is almost invariably due to endocrine disease or a drug effect. Nevertheless, the risk of cancer is much greater in men with gynecomastia; in such men, gross asymmetry of the breasts should arouse suspicion of cancer.

Approximately 90% of male breast cancers contain ERs, and the disease behaves similarly to that in a postmenopausal woman. When matched to female breast cancer by age and stage, its overall prognosis is identical. Male breast cancer is best managed by mastectomy and axillary lymph node dissection or SLNB, although some men prefer breast-conserving therapy. Patients with locally advanced disease or positive nodes should also be treated with irradiation. No randomized studies have evaluated adjuvant therapy for male breast cancer, but extrapolation from treatment with women suggests it is indicated. If the cancer is ER positive, which is often the case, tamoxifen is usually the agent of choice. AIs are also effective in men. Anecdotal evidence supports use of gonadotropin-releasing hormones, such as leuprolide, in combination with an AI, since testosterone is a substrate for the aromatase enzyme. The sites of relapse and spectrum of response to chemotherapeutic drugs are virtually identical for breast cancers in either sex.

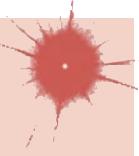
FURTHER READING

- ALLISON KH et al: Estrogen and Progesterone Receptor Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Guideline Update. *Arch Pathol Lab Med* 2020.
- BURSTEIN HJ et al: Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021. *Ann Oncol* 32:1216, 2021.
- CARDOSO F et al: 3rd ESO-ESMO international consensus guidelines for Advanced Breast Cancer (ABC 3). *Breast* 31:244, 2017.
- DHESY-THIND S et al: Use of Adjuvant Bisphosphonates and Other Bone-Modifying Agents in Breast Cancer: A Cancer Care Ontario and American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol* 35:2062, 2017.
- HARRIS LN et al: Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol* 34:1134, 2016.
- PAN K et al: Breast cancer survivorship: state of the science. *Breast Cancer Res Treat* 168:593, 2018.
- QASEEM A et al: Screening for Breast Cancer in Average-Risk Women: A Guidance Statement From the American College of Physicians. *Ann Intern Med* 170:547, 2019.
- SAMADDER NJ et al: Hereditary Cancer Syndromes-A Primer on Diagnosis and Management: Part 1: Breast-Ovarian Cancer Syndromes. *Mayo Clin Proc* 94:1084, 2019.
- SCHMID P et al: Pembrolizumab for early triple-negative breast cancer. *N Engl J Med* 382:810, 2020.
- TUTT ANJ et al: Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer. *N Engl J Med* 384:2394, 2021.
- VISVANATHAN K et al: Use of Endocrine Therapy for Breast Cancer Risk Reduction: ASCO Clinical Practice Guideline Update. *J Clin Oncol* JCO1901472, 2019.
- WOLFF AC et al: Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol* 36:2105, 2018.

80

Upper Gastrointestinal Tract Cancers

David Kelsen



Cancers of the upper gastrointestinal tract include malignancies of the esophagus, stomach, and small bowel. Esophageal, gastroesophageal junction, and gastric cancers are among the most common of human malignancies, with 1.5 million global new cases diagnosed in 2018. In the United States, a lower risk area, it is estimated that in 2020, esophageal cancer will be diagnosed in 18,440 people and cause 16,170 deaths; for gastric cancer, 27,600 new cases will be diagnosed and 11,010 deaths will occur. Small intestine cancers are rare.

ESOPHAGEAL CANCER

■ INCIDENCE AND CAUSATIVE FACTORS

Two distinct forms of cancer with different epidemiologies, causative factors, and genomic profiles arise within the esophagus: squamous cell cancers, which occur more frequent in the upper and mid esophagus; and adenocarcinomas, which are almost always located in the lower esophagus and at the gastroesophageal junction. The incidence of esophageal cancer varies up to 20-fold based on geographic distribution: it is relatively uncommon in North America, but has a high incidence in Asia (especially China), the Normandy coast of France, and Middle Eastern countries such as Iran. This marked global variation is likely due to different causative factors in the development of the malignancy, leading to two different cancer types within the same tissue: squamous cell cancers are more common in high-incidence areas, usually with lower Human Development Index (HDI) scores (a measure of economic development that includes standard of living, health, and education). Overall, approximately 572,000 new cases of esophageal cancer were diagnosed globally in 2018; esophageal cancer was the seventh most common cause of malignancy and the third most common cause of cancer-related mortality, with an estimated 508,000 deaths.

The clearest high-risk factors for the squamous cell cancer subtype in Western countries are alcohol and tobacco abuse; concurrent alcohol and tobacco abuse further increases the risk. Ingestion of extremely hot substances (such as tea in Iran and mate [maté] in South America) has been proposed as a risk factor; in India, chewing the areca (betel) nut increases the risk of esophageal squamous cell cancers. Less common risk factors include chronic achalasia, radiation therapy (such as is delivered for treatment of Hodgkin's lymphoma or breast cancer), lye ingestion, and Plummer-Vinson (Patterson-Kelly) syndrome (iron deficiency anemia, glossitis, cheilosis, and the development of esophageal webs) (**Table 80-1**). Adenocarcinoma of the lower esophagus and gastroesophageal junction has been the predominant histologic subtype in the United States and Western Europe for several decades, now making up >75% of all incident cases. Risk factors for adenocarcinoma (**Table 80-2**) include chronic reflux esophagitis leading to inflammation and the development of Barrett's esophagus (the finding of glandular gastric type mucosa extending into the esophagus). Although obesity increases the risk of reflux esophagitis, a substantial number of patients with newly diagnosed adenocarcinoma of the esophagus and gastroesophageal junction are younger and fit; Barrett's esophagus may still be found in these patients. In patients with adenocarcinoma of the lower esophagus in which Barrett's esophagus is not present, the disease may arise without Barrett's esophagus, or an extensive tumor found at diagnosis may obliterate previous areas of Barrett's. Genomic alterations may be identified even before the development of frank adenocarcinoma in patients with dysplasia associated with Barrett's esophagus. These include mutations of *TP53*, a gene critical in regulating uncontrolled cell division, and aneuploidy in dysplastic regions. Risk of progression of Barrett's esophagus to cancer is about 0.4–0.5% per year. Management of Barrett's esophagus is discussed in **Chap. 323**.

TABLE 80-1 Some Etiologic Factors Associated with Squamous Cell Cancer of the Esophagus

Excess alcohol consumption
Cigarette smoking
Other ingested carcinogens
Nitrates (converted to nitrites)
Smoked opiates
Fungal toxins in pickled vegetables
Mucosal damage from physical agents
Hot tea
Lye ingestion
Radiation-induced strictures
Chronic achalasia
Host susceptibility
Esophageal web with glossitis and iron deficiency (i.e., Plummer-Vinson or Paterson-Kelly syndrome)
Congenital hyperkeratosis and pitting of the palms and soles (i.e., tylosis palmaris et plantaris)
? Dietary deficiencies of selenium, molybdenum, zinc, and vitamin A

As opposed to other gastrointestinal malignancies, such as colorectal cancer, inherited cancer susceptibility genes are rarely associated with esophagus and gastroesophageal junction cancers. An exception is the rare inherited cancer susceptibility gene driving tylosis palmaris and plantaris; a mutation in the *RHBDF2* gene is associated with an increased risk for squamous cell cancers of the esophagus. Lynch syndrome modestly increases the risk of gastric and potentially gastroesophageal junction adenocarcinomas.

■ SCREENING AND SURVEILLANCE OF HIGHER RISK GROUPS

Because of its low incidence in North America and the absence of proven blood-based biomarker for esophageal cancer assays, screening of the asymptomatic general population using, e.g., upper endoscopy is not currently recommended in the United States. Periodic endoscopy is used for surveillance of higher risk patients, such as those with Barrett's esophagus and especially with dysplasia, based on expert opinion guidelines.

■ GENOMIC ALTERATIONS

Within a tissue, subtyping has revealed substantial genomic differences between adenocarcinomas and squamous cell cancers of the esophagus. An integrated analysis involving several different genomic platforms performed by The Cancer Genome Atlas (TCGA) Research Network investigators demonstrated that esophageal squamous cell cancers more closely resembled squamous cell carcinomas of other primary sites, such as the head and neck, than adenocarcinomas arising in the esophagus. Three molecular subclasses of squamous cell cancer were identified (of note, as opposed to squamous cell cancer of the head and neck, human papillomavirus was not identified in any of the three subgroups). Among other differences, the spectrum of genomic amplifications in squamous cell cancers are substantially different than that of adenocarcinomas. In adenocarcinomas, *ERBB2* (*HER2*) was frequently amplified, as were *VEGFA* and *GATA4/6*. The genomic profile for esophageal and gastroesophageal junction adenocarcinomas was very similar to the chromosomally unstable variant of gastric

adenocarcinoma, suggesting that proximal gastric and gastroesophageal junction tumors may have a similar driving factor (see below). Other studies comparing transcriptomes of adenocarcinomas and squamous cell cancers across tissues (i.e., the same tumor histology arising in different organs, such as squamous cell cancers and adenocarcinomas from the esophagus, lung, and uterine cervix) found that histologies among the different organs showed more similarity than between the different histologies within the same organ. In addition to implications regarding driving factors in the initiation and progression of cancer, these genomic alterations are important for therapeutic decisions involving systemic agents given in the neoadjuvant or postoperative adjuvant setting or for advanced metastatic disease. For esophageal cancer, genomic abnormalities that should be considered in prescribing drug-based therapy include analysis for *HER2* amplification, PD-L1 expression, and hypermutated tumors\microsatellite instability (see below).

CLINICAL FEATURES

■ PRESENTING SYMPTOMS

The most common symptoms leading to suspicion of esophageal cancer are dysphagia or odynophagia and, less frequently, hematemesis or melena. More subtle symptoms include anorexia and weight loss, and fatigue and shortness of breath if anemia from gastrointestinal bleeding is present. Because the symptoms of dysphagia or odynophagia are usually not perceived by the patient until substantial obstruction of the esophageal lumen has occurred, the large majority of patients with esophageal cancer are found with locally advanced if not metastatic disease. Patients with symptoms of dysphagia and/or odynophagia should undergo upper endoscopy (rather than a barium contrast study) to determine the presence or absence of malignancy; biopsy should be performed at the same setting to determine histology. Depending on the tumor stage, molecular diagnostic or next-generation sequencing (NGS) analysis to assist in determine potential therapies would be performed. These studies should be done on all patients with metastatic disease as it will guide therapy. NGS requires adequate tumor cellularity, which is sometimes difficult to achieve from endoscopic biopsy. Some high-volume U.S. centers routinely perform NGS on all specimens, including from the primary tumor for patients without metastatic disease.

■ STAGING

Therapeutic strategy is based on the stage of the disease using a system such as the eighth edition of the American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging system. The T stage is based on the size of the tumor and depth of penetration through the esophageal wall (which for most of its course is not covered by serosa so that invasion through the muscle layer leads directly into periesophageal tissues) (Fig. 80-1). Patients with regional lymph node metastases are still potentially curable. Metastatic disease is generally treated with palliative intent with rare exceptions. Because neoadjuvant (preoperative) therapy is widely employed for esophageal cancer in an effort to improve subsequent surgical outcomes, the AJCC TNM staging system includes clinical, pathologic (for patients undergoing initial surgery as first treatment), and ypTNM staging assessment for those treated with preoperative therapy. See Table 80-3 for the TNM staging classification for gastric cancer, which is similar to esophageal cancer.

Determining tumor extent includes careful physical examination, which may reveal palpable lymphadenopathy or hepatomegaly; imaging studies including computed tomography (CT) and fluorodeoxyglucose (FDG) positron emission tomography (PET)/CT scan are used to assess for metastatic disease. If no metastatic disease is identified, endoscopic ultrasonography (EUS) is commonly performed to more definitively determine depth of penetration of the primary tumor (T) and regional lymph node involvement. For tumors of the mid and upper esophagus (5% of esophageal cancers are in the upper third of the esophagus, 20% in the middle third, and 75% in the lower third), bronchoscopy may be performed to rule out invasion of the tracheobronchial tree.

TABLE 80-2 Some Etiologic Factors Associated with Adenocarcinoma of the Esophagus

Chronic gastroesophageal reflux
Obesity
Barrett's esophagus
Male sex
Cigarette smoking

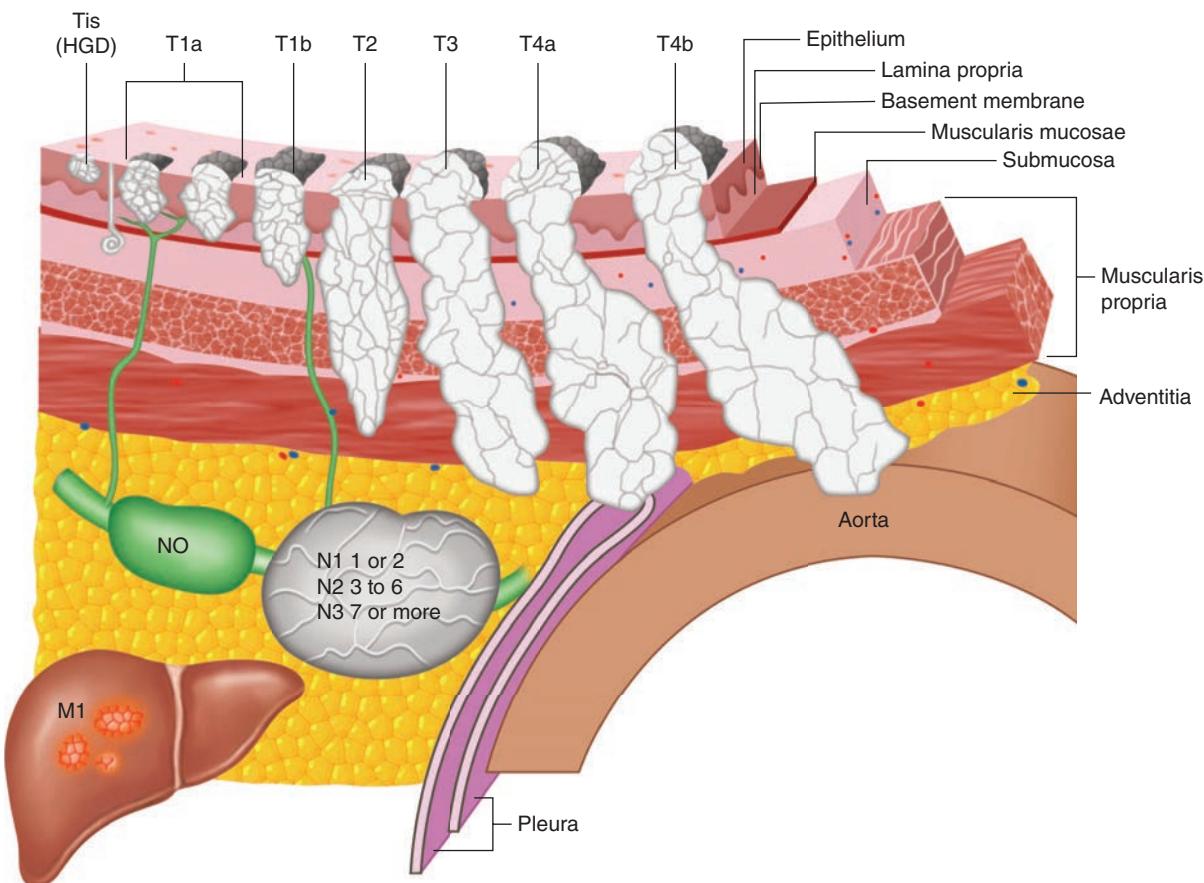


FIGURE 80-1 Patterns of spread of esophageal cancer and the basis for anatomic staging. HGD, high-grade dysplasia. (Reproduced with permissions from TW Rice et al: Cancer of the esophagus and esophagogastric junction: An eighth edition staging primer. *J Thorac Oncol* 12:36, 2017.)

TABLE 80-3 AJCC Prognostic Stage Groups for Esophageal Cancer Using cTNM (Pretreatment)

TNM	CLINICAL STAGE	PRESENTING AT THIS STAGE ^{a,b}	5-YEAR SURVIVAL RATE	
			SQUAMOUS	ADENOCARCINOMA
cTis, N0, M0	0	1.2%	75%	82%
cT1-2, N0, M0	I	17%	75%	78%
cT1-2, N1-3, M0	IIA	7%	53%	50%
cT3-4a, N0, M0	IIB	13%	40%	40%
cT3-4a, N1-3, M0 ^c	III	31%	25%	25%
cT4b, any N, M0	IVA		17%	21%
cAny T, any N, M1	IVB	5%	10%	18%

Survival by ypTNM Staging After Neoadjuvant Chemotherapy

TNM	yp STAGE	ESTIMATED 5-YEAR SURVIVAL RATE	
		SQUAMOUS	ADENOCARCINOMA
T1-2, N0, M0	I	46%	52%
T1, N1, M0			
T3, N0-1, M0	II	34%	38%
T2, N1-2 M0			
T1, N2-3, M0			
T4a, N0, M0			
T4a, N1-3 M0	III	22%	27%
T4b, any N, M0			
T3, N2-3, M0			
T2, N3, M0			
Any T, any N, M1	IV	10%	12%

^aSquamous cell and adenocarcinoma histologies combined. ^bSurgical series; underestimates incidence of M1 disease at presentation. ^cIncidence includes cT4b and cNanyM0.

Sources: Adapted from TW Rice et al: CA Cancer J Clin 67:304, 2017; TW Rice et al: Dis Esophagus 29:707, 2016; and TW Rice et al: personal communication.

The finding of invasion of the trachea or bronchus rules out surgical intervention with curative intent. Regional lymph nodes may be biopsied under EUS guidance. If metastatic disease is suspected, biopsy to confirm tumor staging and to obtain adequate tissue for molecular and genomic alterations analysis should be performed. If systemic therapy is indicated as a portion of the treatment (for metastatic disease or for preoperative therapy for locally advanced cancers), serial FDG-PET/CT scans, using decrease in FDG avidity as a surrogate measure of effectiveness, are increasingly being used to guide whether the initial therapy should be continued or changed.

TREATMENT

Esophageal Cancer

Although the prognosis for patients with esophageal cancer (all stages) is still poor, a slow but steady improvement in 5-year survival has been noted. Because no effective early detection methods exist, the number of patients found to have very-early-stage cancers at the time of diagnosis has not markedly increased; the modest improvement in survival is probably a combination of somewhat improved systemic therapy as well as decreased operative morbidity and mortality when surgery is performed by high-volume surgeons at high-volume centers, as well as improvements in the delivery of external-beam radiation therapy.

For patients without evidence of metastatic disease, the goal of therapy is cure, usually by employing combined-modality therapies. Except for patients with early-stage esophageal cancer, which might be treated by surgery alone (or for very-early-stage lesions, by endomucosal resection alone), systemic drug therapy plus external-beam radiation therapy is a standard of care option for esophageal cancers. For selected patients with gastroesophageal cancers, systemic therapy alone may be given before definitive surgical resection. For patients with squamous cell cancers of the upper and mid esophagus, combined chemotherapy plus concurrent radiation therapy is a standard of care option, with surgery reserved for patients not achieving a complete radiographic and endoscopic response. Chemotherapy plus concurrent radiation was superior to radiation therapy alone in several clinical trials. Increasingly, all systemic therapy given with curative intent is given before operation, although if surgery is the initial therapy and the patient is found to have more locally advanced cancer at pathology (e.g., regional lymph node metastasis), postoperative systemic therapy is used in the adjuvant setting. Adjuvant chemotherapy is more frequently indicated in patients with adenocarcinoma than squamous cell cancers.

For patients with metastatic disease, the goal of therapy is symptom palliation and life extension. No randomized trials of supportive care only versus systemic therapy plus best supportive care have been reported in patients with esophageal cancers. For gastric cancer (a similar histology as distal esophageal and gastroesophageal junction tumors as discussed above), clinical trials performed in the 1980s and 1990s indicated a modest improvement in 1- and 2-year survival when systemic therapy was initiated versus best supportive care only. While the cytotoxic chemotherapy regimens used for palliation have not changed dramatically over the past 10 years, subgroups of patients have been identified who benefit from therapies targeting specific genomic alterations. Approximately 20–25% of patients with adenocarcinoma of the esophagus or gastroesophageal junction are found to have amplified or overexpressed HER2; trastuzumab plus chemotherapy results in higher response rates and longer progression-free and overall survival compared to chemotherapy alone. Immune modulation therapy using PD-1 inhibitors is second-line palliative therapy for patients who have esophageal cancers expressing PD-L1 or having hypermutated or microsatellite-unstable genotype. Molecular diagnostic or genomic alteration analysis assays to identify these biomarkers should be performed routinely in patients with metastatic esophageal cancer to help guide therapy.

Supportive measures to improve nutrition and quality of life include placement of an endoluminal stent in the setting of high-grade obstruction; use of enteral nutrition can also be performed using a percutaneous gastrostomy. Photodynamic therapy and endoscopic laser therapy have been used to treat endoluminal obstruction.

TUMORS OF THE STOMACH

■ ADENOCARCINOMA OF THE STOMACH

Incidence and Causative Factors A century ago, gastric adenocarcinomas were among the most common of malignancies in the United States. Since the 1920s, the incidence of gastric cancer has steadily decreased; while the reason for this has not been definitively identified, it coincided with widespread use of refrigeration and a decreased need for food preservatives. In 2020, it is estimated that there will be 27,600 new cases of gastric cancer diagnosed in the United States; while now seen much less frequently, it remains a lethal disease, with 11,010 deaths. Globally, gastric cancer is still very common, with an overall global incidence of 1.03 million new cases per year and 780,000 deaths, making gastric cancer the third most common cause of cancer mortality. High-incidence areas, as is the case for esophageal cancers, include large Asian countries such as China, Korea, and Japan; South American countries such as Chile; and Eastern European countries.

While the number of new cases of body and distal gastric cancers has decreased in Western, high-HDI countries, the incidence of adenocarcinomas of the gastroesophageal junction has markedly increased in the same areas over the past several decades. The ingestion of high concentrations of nitrates found in dried, smoked, and salted foods may be a contributing factor. Bacteria such as *Helicobacter pylori* and ingestion of partially decayed bacterially contaminated food may lead to the generation of carcinogenic nitrites from nitrates (Table 80-4). A causative factor is suspected to be chronic inflammation due to reflux of gastric contents into the esophagus, particularly in obese people. Obesity alone is not the cause, as a substantial number of these patients are fit and not overweight. Early-onset gastric cancers (gastric cancer occurring in patients under the age of 50), primarily proximal or gastroesophageal junction cancers, have also increased. A second cause of chronic inflammation, *H. pylori* infection, is a known driver in many cases of gastric cancer. While *H. pylori* is extremely common, occurring in approximately half of all humans, gastric cancer occurs in only a small subset of those infected. Higher cancer risk has been associated with certain strains of *H. pylori*; a specific human genetic predisposition has not yet been identified. Supportive evidence that *H. pylori* infection is a causative factor in the development of gastric cancer includes prospective studies demonstrating that treatment of *H. pylori* infection decreases the overall risk of gastric cancer. For example, patients with *H. pylori* infection who had at least one first-degree relative with a history of gastric cancer (increasing their own risk of stomach cancer) were randomly assigned to placebo or treatment for *H. pylori*. The group receiving *H. pylori* eradication showed a significant decrease in the incidence of gastric cancer (especially for

TABLE 80-4 Nitrate-Converting Bacteria as a Factor in the Causation of Gastric Carcinoma^a

Exogenous sources of nitrate-converting bacteria:

Bacterially contaminated food (common in lower socioeconomic classes, who have a higher incidence of the disease; diminished by improved food preservation and refrigeration)

Helicobacter pylori infection

Endogenous factors favoring growth of nitrate-converting bacteria in the stomach:

Decreased gastric acidity

Prior gastric surgery (antrectomy) (15- to 20-year latency period)

Atrophic gastritis and/or pernicious anemia

? Prolonged exposure to histamine H₂-receptor antagonists

^aHypothesis: Dietary nitrates are converted to carcinogenic nitrites by bacteria.

those in whom *H. pylori* was successfully eradicated) compared to the control group. Earlier studies had demonstrated that treatment of *H. pylori* in Korean patients who had a prior very-early-stage gastric cancer decreased the incidence of a second gastric cancer. These data suggest that treatment of asymptomatic *H. pylori* gastric infection should be considered for patients who have a first-degree relative who has had gastric cancer or who themselves have a prior history of an early-stage gastric cancer.

In addition to chronic inflammatory conditions, inherited cancer susceptibility genes increase the risk of gastric cancer. These include mutations of *CDH1*, which encodes for the cell cohesion gene e-cadherin; germline *CDH1* mutations markedly increase the risk for the diffuse cell (signet cell) gastric cancer subtype (see below for discussion of histologic subtypes). Patients with an inherited deleterious *CDH1* mutation are considered for prophylactic gastrectomy. *CDH1* mutations also increase the risk for lobular breast cancer. Germline mutations in the mismatch repair pathway (Lynch syndrome) slightly increase the risk for gastric cancer. Other inherited cancer susceptibility genetic syndromes that increase the risk of gastric cancer include familial adenomatous polyposis, juvenile polyposis, and Peutz-Jeghers syndrome. Inherited cancer susceptibility genes such as *BRCA* mutations may not significantly increase risk for gastric cancer. Surveillance programs for the higher risk germline cancer susceptibility genes should be employed.

Gastric cancer stem cells originating in the bone marrow may play an important role in the development of gastric cancer. *H. pylori* may be an inciting factor for recruitment of such bone marrow gastric stem cells. If this hypothesis is confirmed, it may have important implications for therapy of gastric cancers.

Clinical Features • SURVEILLANCE STRATEGIES As is the case for esophageal cancer, the overwhelming majority of Western patients with gastric cancer are symptomatic at the time of diagnosis. Early detection programs, in Japan and Korea, where gastric cancer has been among the most common of malignancies (although its incidence has been decreasing), include upper endoscopy and, in Japan, upper endoscopy and serum pepsinogen; these programs have increased the number of patients found with early gastric cancer and decreased mortality rates. This strategy has not been cost effective in populations in which the incidence of gastric cancer is much lower, such as in the United States. In high-incidence areas, treatment of symptomatic *H. pylori* is a preventive measure. Exceptions to routine surveillance in the United States are asymptomatic patients with *CDH1* (and other cancer susceptibility gene) mutations who may be part of early detection programs and in whom prophylactic gastrectomy is a management option.

PRESENTING SYMPTOMS Presenting symptoms include vague upper abdominal discomfort, hematemesis or melena, anorexia and early satiety, and unexplained weight loss. For patients with esophagogastric junction cancers, dysphagia or odynophagia may be the presenting symptom. Anemia may be found due to occult bleeding. These symptoms and signs lead to upper (and if site of bleeding is uncertain, lower) endoscopy and biopsy (endoscopy has long replaced barium contrast radiography as an initial diagnostic step). Occasionally, imaging using CT performed to evaluate abdominal symptoms may identify gastric thickening or a gastric mass leading to upper endoscopy. Physical examination can reveal left supraclavicular adenopathy (Virchow's node), a periumbilical mass (Sister

Mary Joseph nodule), a pelvic mass on rectal exam (Blumer's shelf), ascites, or an ovarian mass (Krukenberg tumor). More commonly, physical examination is unrevealing.

Upper endoscopy may reveal an ulcer or ulcerated mass, biopsy of which shows adenocarcinoma. For the diffuse subtype of gastric cancer, a mass or ulceration may not be seen, but rather, thickened gastric rugae may be noted. Initial biopsy may not reveal diffuse gastric cancer, which may track below the mucosal surface. In these patients, EUS may guide biopsy.

Histopathology Classification of Primary Gastric Adenocarcinomas The large majority (~85%) of gastric malignancies are adenocarcinomas or subtypes of adenocarcinoma. Other malignancies, discussed below, include neuroendocrine tumors (carcinoid tumors), primary gastric lymphomas, gastrointestinal stromal tumors (GISTs), and other rare malignancies. Using the Lauren classification, pathologists classify adenocarcinomas on the basis of histopathology as intestinal (more common) or diffuse subtype (~20%). As noted above, the diffuse subtype is associated with inherited *CDH1* mutations; in addition, in the TCGA genomic analysis of gastric cancer, approximately a third of diffuse subtype cases had somatic *CDH1* mutations. The intestinal subtype is associated with *H. pylori* infection and atopic gastritis. Histologic grade also influences the clinical course.

Genomic analysis performed by several groups has resulted in molecular classifications of gastric cancer that may, in the future, inform staging systems, provide a better understanding of the driving factors in the development of gastric cancer, and provide important information on treatment options (Fig. 80-2). For example, the TCGA group reported the results of a multiplatform analysis of 295 patients with previously untreated gastric cancer; both Western and Asian patients were included in the analysis. Four subtypes of gastric cancer were identified: high Epstein-Barr virus burden, microsatellite unstable with hypermutation, genetically stable (associated with the diffuse subtype), and chromosomal unstable. The Asian Cancer Research Group (ACRG), studying primary tumors from 300 Korean patients, analyzed gene expression profiles and found four subtypes: mesenchymal, microsatellite unstable, microsatellite stable with *TP53* expressed, and microsatellite unstable with *TP53* mutated. Clinical outcome was correlated with genomic subtype in both studies, with microsatellite

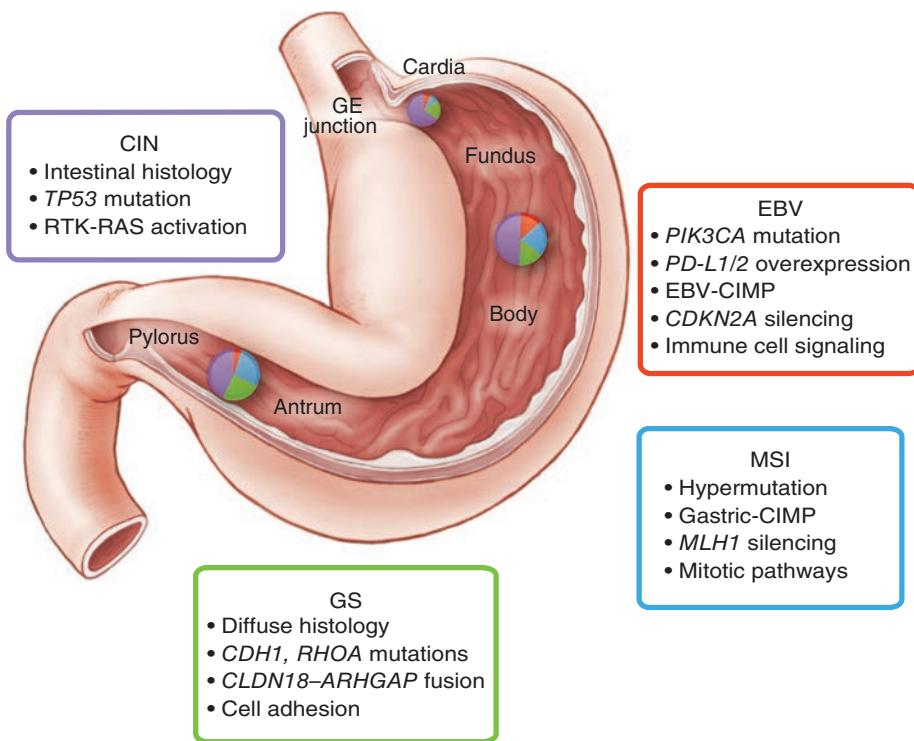


FIGURE 80-2 Molecular/genomic characterization of subtypes of gastric carcinomas. CIMP, CpG-island methylator phenotype; CIN, chromosomally unstable; EBV, Epstein-Barr virus-associated; GE, gastroesophageal; GS, genetically stable; MSI, microsatellite instability-associated.

unstable tumors having the best outcome and genetically stable (TCGA) and mesenchymal (ACRG) types the worst.

In addition to histopathology, molecular diagnostics (including NGS) are an important part of the pathology workup. The molecular subtypes have therapeutic implications; for example, ~20% of gastric cancer or gastroesophageal junction cancer patients' tumors have overexpression or amplification of *HER2*, which would lead to the addition of agents such as trastuzumab as part of systemic treatment for metastatic disease. Immune modulation therapy may be used in patients with hypermutated tumors, found by NGS or by polymerase chain reaction (PCR) for microsatellite instability (MSI). An evaluation for overexpression or amplification of *HER2*, quantification of PD-L1 by immunohistochemistry, and assessment of MSI by PCR or deficient mismatch repair protein (dMMRP) expression should be a routine part of the pathology workup of patients with metastatic gastric cancer.

More controversial is whether these assays should also be routinely performed in patients with potentially operable gastric cancer because the addition of trastuzumab to neoadjuvant chemotherapy has not yet been shown to change outcome. In large-volume centers, NGS is routinely performed on pretreatment biopsies. Currently, the finding on pathologic assays of positive tumor Epstein-Barr virus (identified in 8–10% of gastric cancer patients) does not change therapeutic options.

Staging Once a diagnosis of a primary gastric adenocarcinoma is made, algorithms for clinical evaluation of stage include physical examination and imaging studies (Fig. 80-3; Table 80-5). Tumor-related biomarkers such as carcinoembryonic antigen (CEA) or CA19-9 may be elevated but are nonspecific (may be elevated in a number of other gastrointestinal and other site cancers). Diagnostic CT scan of the chest, abdomen, and pelvis should be performed. If metastatic

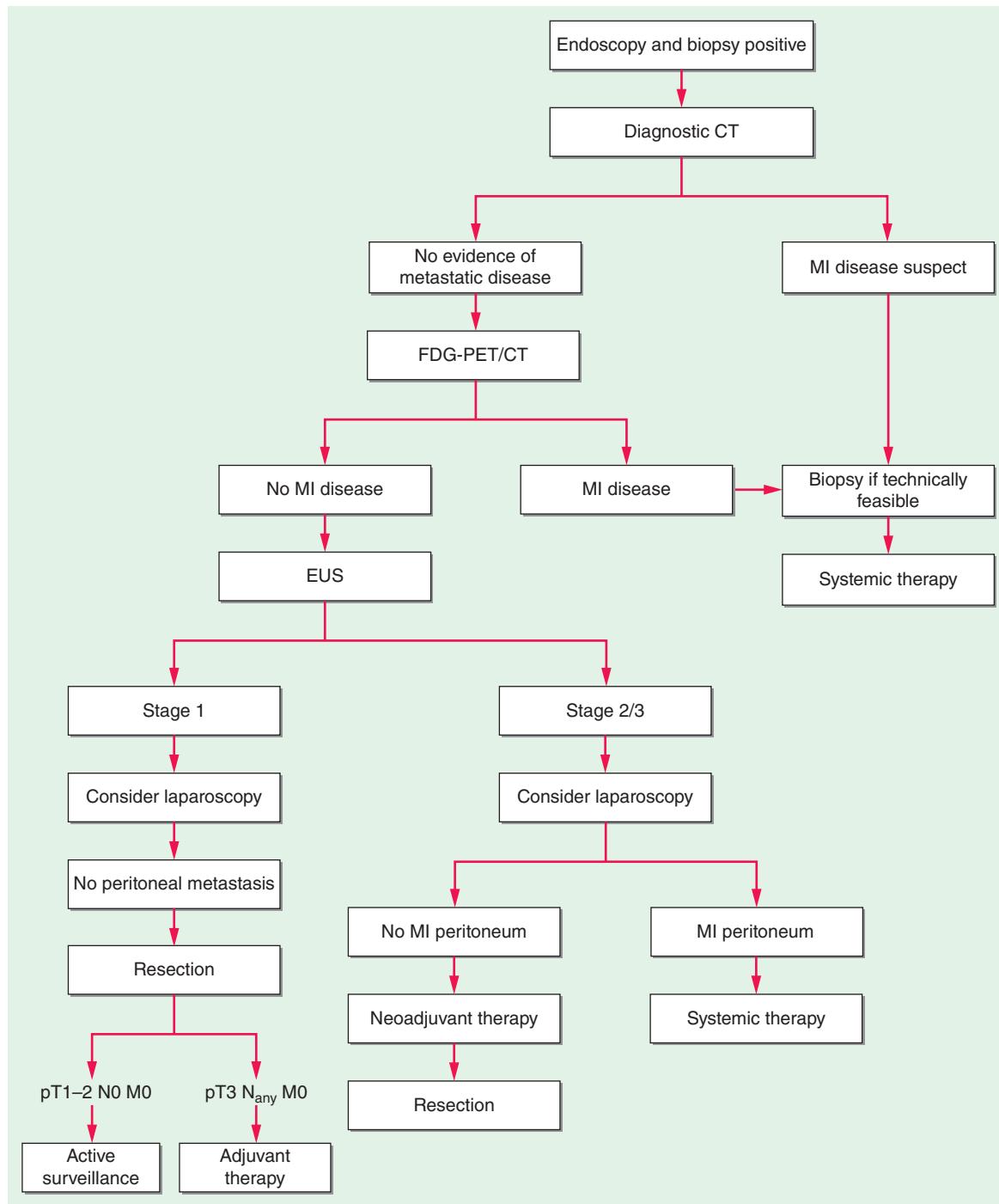


FIGURE 80-3 Staging for gastric adenocarcinoma. CT, computed tomography; EUS, endoscopic ultrasound; FDG-PET, fluorodeoxyglucose positron emission tomography.

TABLE 80-5 Staging System for Gastric Carcinoma

STAGE	TNM	FEATURES	DATA FROM ACS IN THE UNITED STATES	
			NO. OF CASES, %	5-YEAR SURVIVAL, %
0	TisNOMO	Node negative; limited to mucosa	1	90
IA	T1N0MO	Node negative; invasion of lamina propria or submucosa	7	59
IB	T2N0MO T1N1MO	Node negative; invasion of muscularis propria	10	44
II	T1N2MO T2N1MO T3N0MO	Node positive; invasion beyond mucosa but within wall <i>or</i> Node negative; extension through wall	17	29
IIIA	T2N2MO T3N1-2MO	Node positive; invasion of muscularis propria or through wall	21	15
IIIB	T4N0-1MO	Node negative; adherence to surrounding tissue	14	9
IIIC	T4N2-3MO	>3 nodes positive; invasion of serosa or adjacent structures		
	T3N3MO	7 or more positive nodes; penetrates wall without invading serosa or adjacent structures		
IV	T4N2MO T1-4N0-2M1	Node positive; adherence to surrounding tissue <i>or</i> Distant metastases	30	3

Abbreviations: ACS, American Cancer Society; TNM, tumor-node-metastasis.

disease is suspected on imaging, a biopsy of a metastatic site should be strongly considered to confirm stage IV disease, which changes the goals of care from potentially curative to palliative treatment. As is the case for esophageal cancer, FDG-PET/CT, which is more sensitive than diagnostic CT scan in identifying sites of metastatic disease, should be performed if the anatomic CT is negative for metastatic disease. Note, however, that FDG-PET may be noninformative (the primary tumor may not be FDG-avid, particularly in diffuse-type gastric cancer). If imaging does not reveal metastatic disease, EUS should be strongly considered to determine depth of penetration of the primary tumor and the presence or absence of regional lymphadenopathy suspicious for metastasis. Lymph node biopsy and, on occasion, biopsy of left hepatic parenchymal lesions found on EUS can be performed at the same setting. Endoscopic biopsies usually provide enough tumor tissue for molecular diagnostic pathology testing for HER2, MSI/MMRP, and PD-L1 assessment; it may not provide enough tissue for genomic alteration analysis. If neoadjuvant therapy is planned, laparoscopy should be considered to allow evaluation of the peritoneal cavity, with peritoneal washing for cytology if no peritoneal metastases are visible. The peritoneal cavity is a common site of metastases, especially from diffuse-type gastric cancer. The finding of peritoneal involvement either visibly or by positive cytology is staged as metastatic disease and diminishes the chance for curative resection.

The staging classification for gastric cancer is summarized in Table 80-5. Three staging classifications are provided: cTNM clinical staging (before any therapy has been given), pTNM pathologic staging (for patients not undergoing preoperative therapy), and a post-neoadjuvant therapy classification staging (ypTNM). The three components take

into account current standard of care options for therapy in which the AJCC prognostic stage groups from clinical staging guide therapeutic decisions. For example, after clinical evaluation, a large percentage of newly diagnosed patients will be found to have higher-stage primary cancers (penetrating through the gastric wall [T3 or T4] or lymph node-positive tumors), in which case perioperative (neoadjuvant) systemic therapy may be chosen. Pathologic examination of the resected specimen for prognostic stage classification must take into account exposure to preoperative therapies that may lead to downstaging (thus, ypTNM staging). Nomograms have been developed for predicting outcome in patients undergoing surgery as initial treatment.

TREATMENT

Gastric Cancer

POTENTIALLY CURABLE GASTRIC CANCER: SURGERY

Surgical removal of the primary tumor with negative microscopic margins (an R0 resection) and with resection of regional lymph nodes is currently the only curative therapy; with surgery alone, 5-year survival rates are approximately 25%. If tumor cells are found at the margin of resection (R1) or if visible cancer is left at the time of surgical removal of the primary tumor (R2), surgery is palliative rather than curative. For patients with early-stage tumors (mostly clinical stage I), surgery without perioperative systemic therapy may be performed. For patients with more locally advanced tumors (clinical stages II A, II B, III), who compose approximately 70% of newly diagnosed operable patients, multimodality therapy (surgery and systemic chemotherapy) improves overall survival. Both neoadjuvant (preoperative) and postoperative systemic therapy are accepted approaches. If staging studies demonstrate a locally advanced cancer (T3/4 or node positive), preoperative treatment is recommended in a medically fit patient. If surgery is performed first and a locally advanced cancer is found, postoperative chemotherapy or chemotherapy plus chemoradiation is recommended. For selected very-early-stage gastric cancers (primary tumors that are ≤2 cm in diameter, are well to moderately differentiated, do not invade the deep submucosa, and do not show lymphovascular invasion or lymph node metastasis), which are not commonly found in the United States, endoscopic mucosal resection (EMR) or endoscopic submucosal dissection (ESD) may be performed by experienced gastroenterologists in place of surgical resection, with favorable results in studies in high-incidence areas such as Japan.

For patients in whom the primary tumor is in the distal stomach, a subtotal gastrectomy is the preferred surgical procedure. For tumors of the proximal stomach, the options for resection include total gastrectomy or, alternatively, proximal gastrectomy. Esophagogastrectomy is performed for tumors involving the gastroesophageal junction. In selected patients, a jejunostomy feeding tube may be placed if postoperative radiation therapy is part of the treatment plan.

As noted above, laparoscopy is commonly performed at high-volume centers before a final decision regarding the role of surgery. If staging has already demonstrated clinically suspicious lymph nodes or an advanced T stage tumor, but laparoscopy does not demonstrate peritoneal metastasis, perioperative chemotherapy is given before surgical resection.

Palliative resection of the primary tumor is usually performed only if symptoms such as uncontrollable bleeding or obstruction are present that cannot be relieved by other means.

As is the case for colorectal cancer, a correlation exists between the number of lymph nodes removed and sampled and outcome. Sentinel lymph node biopsy is not performed in gastric cancer outside of a research study setting. The goal is to examine at least 15 lymph nodes from the resected specimen; it is more controversial whether more extensive lymph node resection itself affects outcome; the extent of lymphadenopathy can be classified using a D0-D3 system with a higher number meaning more extensive lymphadenopathy. In the United States, a modified D2 (D1+) is

resection preserving the spleen and avoiding pancreatectomy is recommended but should be performed by experienced surgeons at high-volume centers. Japanese investigators and others have used very extensive lymph node dissections, but studies have not demonstrated an advantage for a D3 resection. Both resection of the primary tumor and its regional lymph nodes can be performed laparoscopically in appropriate patients.

In the hands of experienced surgeons, operative mortality would be anticipated to be $\leq 2\%$.

NEOADJUVANT AND POSTOPERATIVE ADJUVANT THERAPY FOR RESECTABLE GASTRIC CANCER

The large majority of potentially resectable Western gastric cancer patients have locally advanced tumors (cTNM stage IIA/B or III). Multimodality therapy using systemic chemotherapy improves 5-year survival rates by 10–15% compared to surgery alone. A widely cited study, the MAGIC clinical trial, randomly assigned patients with potentially resectable disease to receive perioperative chemotherapy or to proceed directly to surgery. Five-year overall survival for patients undergoing surgery alone was 23%; for those receiving pre- and postoperative chemotherapy, it was 36%. On the basis of this and other clinical trials, for most medically fit patients with stage cTNM II and III resectable gastric cancers, preoperative systemic chemotherapy followed by resection and, if tolerable, postoperative chemotherapy is a standard of care. Chemoradiation as given for esophageal cancers is usually used for gastroesophageal junction tumors. Preoperative chemoradiation or preoperative chemotherapy followed by chemoradiation for gastric cancer, as opposed to esophageal or gastroesophageal junction cancers, has been studied but is still an investigational approach. For patients being treated with multimodality therapy, close interactions among the surgeon, medical oncologist, and radiation oncologist are essential.

Clinical trials have compared different cytotoxic chemotherapy regimens, most of which include a platinum compound—either cisplatin or oxaliplatin. Currently, a platinum compound plus a fluorinated pyrimidine, such as fluorouracil or capecitabine, given for three to four cycles before surgery is a standard of care option. Drug combinations are favored; an example is the FOLFOX regimen, which includes fluorouracil, leucovorin, and oxaliplatin. For very fit patients, a combination of fluorouracil, oxaliplatin, and docetaxel (FLOT) may be chosen. Addition of trastuzumab to chemotherapy has not improved outcomes for patients with HER2-positive cancers. Careful monitoring of chemotherapy-related toxicities with appropriate dose modifications is important. Several clinical trials have included both preoperative and postoperative systemic therapy; a substantial number of patients will have a slow postoperative recovery and not receive postoperative treatment. Maximizing the ability to give preoperative chemotherapy is an important consideration. For patients receiving preoperative systemic chemotherapy and undergoing a D2/D1+ dissection, postoperative chemoradiation therapy has not improved outcome.

For patients in whom the primary tumor has been resected and who did not receive preoperative chemotherapy, who are found to have stage II or III cancers, or who have < 15 lymph nodes found in the resected specimen, postoperative chemoradiation is a treatment option. Chemoradiation therapy may also be given for unresectable cancers in selected patients.

PALLIATIVE THERAPY FOR INCURABLE GASTRIC CANCER

Patients with clinical stage IV gastric cancers with an adequate performance status should be offered systemic drug therapies. Small clinical trials performed in the 1980s and 1990s showed a survival benefit for systemic therapy compared to best supportive care only. The cytotoxic chemotherapy regimens most commonly employed are still based on a platinum compound and a fluorinated pyrimidine (e.g., FOLFOX, as is used in the perioperative setting). However, two subgroups of gastric cancer patients have been identified who benefit from the addition of noncytotoxic

agents. Those whose tumors have overexpressed or amplified *HER2* should receive *HER2*-targeted agents such as trastuzumab plus cytotoxic chemotherapy because a modest survival advantage has been demonstrated. Additional *HER2*-targeted therapy using trastuzumab-deruxtecan, a monoclonal antibody-drug conjugate, has encouraging activity in patients whose tumors were refractory to trastuzumab. For patients with MSI/dMMRP gastric cancers, PD-1 inhibitors such as pembrolizumab should be used (currently approved in the second-line setting). Immune modulation therapy using PD-1 inhibitors such as pembrolizumab or nivolumab is also approved in gastric cancer with tumor specimens having $\geq 1\%$ PD-L1 expression, with modest response rates. The development of more effective immune therapies and their combination with cytotoxic chemotherapy (and in combination with trastuzumab plus chemotherapy in *HER2*-positive patients) as initial treatment are areas of active investigation.

When disease progresses after first-line treatment, other therapies include the combination of a VEGF receptor-targeted agent, ramucirumab, either alone or in combination with paclitaxel. As noted above, immune modulation therapy may cause remissions in patients whose tumors have at least 1% PD-L1 expression. PD-1 inhibitors are the preferred option for patients whose tumors are microsatellite unstable. Several other cytotoxic agents have activity in the palliative setting including irinotecan and trifluridine-tipiracil. Best results from clinical trials indicate overall survival for treated patients with stage IV disease is still only 12–15 months.

Radiation therapy using shorter regimens may be employed to palliate bleeding. For patients with advanced incurable disease, other supportive measures include placement of a duodenal stent to relieve gastric outlet obstruction; in selected patients, surgical procedures for gastric outlet obstruction may be performed. Radiation therapy might be used if not previously given. Enteral feeding using a jejunostomy tube may support nutritional needs.

GASTRIC LYMPHOMAS

Lymphomas of the stomach are an uncommon (~3%) but important subgroup of gastric malignancies. They are extranodal non-Hodgkin's lymphomas (NHL). The gastrointestinal tract is the most common site for extranodal NHL, and the stomach is the most common site within the gastrointestinal tract. The presenting symptoms are similar to those of the much more common adenocarcinoma of the stomach described above, including pain, anorexia, and bleeding. Symptoms of fever and night sweats occur in 10–15% of patients with gastric NHL. Because the treatment options are so different, obtaining adequate tissue for definitive pathologic examination is crucial in diagnosing gastric lymphomas. On occasion, this may be challenging because, similar to diffuse subtype adenocarcinoma, lymphomas may track below the mucosal surface. Multiple deep biopsies and mucosal resection may be needed to provide enough tissue for definitive pathologic assessment.

Potential driving forces in the development of gastric lymphomas include active or prior *H. pylori* infection, which is associated with mucosa-associated lymphoid tissue (MALT) subtype gastric lymphomas. MALT lymphomas may develop in nearly any organ, but the stomach is the most frequent primary site, accounting for ~35% of all MALT lymphomas. Antibacterial therapy directed against *H. pylori* can be a highly effective treatment in these patients. Other forms of NHL may involve the stomach either as primary gastric lymphoma or as a secondary site of disease, including both B-cell (e.g., mantle cell lymphoma, Burkitt's lymphoma, and follicular lymphoma) and T-cell lymphomas (e.g., enteropathy-associated T-cell lymphoma, anaplastic large-cell lymphoma, and peripheral T-cell lymphoma).

Staging is performed in a fashion similar to gastric adenocarcinoma, but the staging classification is different (see below). In addition to a contrast-enhanced diagnostic-quality CT scan of the chest, abdomen, and pelvis, an FDG-PET/CT scan may be helpful. EUS may be used to determine depth of invasion in patients in whom no evidence of metastatic disease is noted. Examination of the peripheral blood and bone marrow aspirate should be considered as part of the workup.

In all patients with gastric lymphoma, *H. pylori* infection status should be evaluated. If *H. pylori* testing is negative by histopathology, noninvasive testing by either stool antigen test or urea breath test should be used. If rituximab will be part of the treatment plan (see below), hepatitis B testing should be performed.

■ STAGING

The TNM staging system is not employed for gastric lymphomas. The Lugano staging system for gastrointestinal lymphomas (a modification of the Ann Arbor staging system) divides patient groups into stages I, II, and IV. Stage I tumors are limited to the gastric wall; stage II tumors have regional lymph node involvement or invasion of local structures; stage IV tumors have either more extensive lymph node involvement or have distant metastasis, including to the bone marrow or other extranodal sites.

■ PATHOLOGIC CLASSIFICATION

The two most common histologic subtypes of gastric lymphoma are marginal zone B-cell lymphomas (gastric marginal zone B-cell lymphomas or MALT; ~40% of newly diagnosed patients) and diffuse large B-cell lymphomas (DLBCLs; ~55%). The distinction is critical because therapeutic options are different. As part of the pathologic evaluation, for patients with *H. pylori*-positive gastric lymphomas, the finding of a t(11;18) translocation identifies a subgroup less likely to respond to *H. pylori* eradication. This translocation may be detected using PCR or fluorescent in situ hybridization (FISH); it creates a chimeric protein composed of the amino terminal of *APII* (apoptosis inhibitor) and the carboxy terminus of *MALT1*, leading to activation of nuclear factor- κ B signaling.

TREATMENT

Gastric Lymphoma

Unlike adenocarcinoma of the stomach, surgical resection has no role in the treatment of primary gastric lymphoma in the absence of complications of therapy such as perforation or uncontrollable bleeding. Resection of gastric lymphoma does not improve outcome.

For patients with MALT lymphoma, eradication of *H. pylori* with antibiotics is highly effective therapy. If tests for *H. pylori* are positive and t(11;18) translocation assay is negative, one of the currently accepted antibiotic regimens for treating *H. pylori* should be the initial therapy. *H. pylori* eradication is associated with high response rates including complete remissions in the majority of patients. The time to remission may be prolonged (in some studies averaging 15–16 months); therefore, careful monitoring is important before determining that a MALT tumor is not responding to anti-*H. pylori* therapy. For patients in whom the t(11;18) translocation assay is positive, options for therapy include anti-*H. pylori* antibiotic therapy plus involved-field radiation therapy or, if radiation is contraindicated, the use of single-agent rituximab, a monoclonal antibody targeting CD20. For patients who are *H. pylori* negative, moderate-dose (24–30 Gy) involved-site radiation therapy or single-agent rituximab is a treatment option. For selected, more advanced, stage IV MALT patients who have not responded to or who have progressed after receiving anti-*H. pylori* antibiotic therapy and/or rituximab, cytotoxic chemotherapy regimens such as R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone) or rituximab and lenalidomide may be considered.

DLBCL may be a result of transformation from more indolent MALT lymphoma or may arise de novo. De novo tumors are more likely to be BCL2 and CD10 positive. MALT lymphomas that have transformed to DLBCLs are more frequently BCL2 and CD10 negative.

For patients with DLBCL, earlier stage tumors may be treated by combination chemotherapy alone or chemotherapy plus involved-field radiation therapy. For more advanced gastric DLBCL tumors, chemotherapy using R-CHOP or R-EPOCH (rituximab, etoposide, prednisolone, vincristine, cyclophosphamide, and doxorubicin)

regimens is standard of care. Some reports have suggested that eradication of *H. pylori* is effective treatment for early-stage DLBCL when the patient also has *H. pylori*.

UNCOMMON TUMORS OF THE ESOPHAGUS AND STOMACH

■ NEUROENDOCRINE TUMORS

Neuroendocrine tumors (NETs) of the esophagus are rare, accounting for <1% of gastrointestinal NETs. They may present with dysphagia and odynophagia similar to that of more common squamous cell or adenocarcinoma of the distal esophagus and gastroesophageal junction or with more nonspecific symptoms such as substernal discomfort or burning consistent with reflux esophagitis. A potential driving factor of higher grade NET is smoking. The initial diagnostic evaluation includes upper endoscopy and biopsy. Pathology may reveal a well-differentiated grade 1 or grade 2 NET with a low metastatic potential; at the other end of the spectrum are small-cell or large-cell NETs, which are fully malignant and frequently metastasize. EUS to assess depth of penetration and presence or absence of regional lymph node metastasis is usually performed. Imaging studies include CT or FDG-PET/CT scan to assess for metastatic disease, particularly in higher grade tumors. For lower grade tumors, somatostatin analog imaging studies such as gallium-68 DOTATATE may be performed if metastatic disease is suspected. For lower grade tumors, endoscopic resection including EMR or ESD may be performed. Small-cell or large-cell NETs that are not metastatic are usually treated with chemotherapy plus external-beam radiation therapy using chemotherapy regimens similar to those employed for small- and large-cell neuroendocrine cancers of the lung (Chap. 84). Systemic therapy for metastatic small- and large-cell esophagogastric NETs is also modeled on therapy for small- and large-cell thoracic NETs.

Gastric NETs (also called gastric carcinoid tumors) represent 7–9% of gastrointestinal NETs but <1% of gastric neoplasms. They are divided into three types (Chap. 84). For all gastric NETs, initial evaluation includes upper endoscopy and biopsy. EUS may be helpful in assessing depth of invasion for larger tumors and for assessing regional lymph node metastases in type 3 tumors. Somatostatin analog imaging using gallium-68 DOTATATE PET/CT scanning may be performed if metastatic disease is suspected. The finding of unresectable metastatic disease that is gallium-68 DOTATATE avid not only provides staging information but also guides potential therapy using somatostatin receptor-targeted therapy.

TREATMENT

Gastric Neuroendocrine Tumors

Type 1 tumors are usually treated endoscopically with polypectomy or endomucosal resection. For larger tumors (>2 cm) or tumors invading through the muscularis or to regional lymph nodes, surgical resection is recommended. Type 2 tumors have a higher risk for regional lymph node metastasis and are usually treated surgically, although selected patients may have a combination of endoscopic resection and limited surgical resection. Type 3 tumors are not associated with elevated gastrin levels and have a higher propensity for regional lymph node metastasis and distant metastasis. Surgery is the treatment of choice for localized type 3 tumors, although EMR has been used in selected patients. Adenocarcinoma of the stomach may be found in 5–10% of type 3 tumors.

■ GASTROINTESTINAL STROMAL TUMORS

Gastrointestinal stromal tumors (GISTs) are rare tumors of the gastrointestinal tract associated with somatic mutations in the *cKIT* (the majority) or *PDGFRA* genes, which are both driver mutations and targets for systemic therapy for metastatic disease; in a minority of cases, neither gene is mutated. GISTs arise from Cajal cells, which bridge between the autonomic nerves to the muscle layer of the bowel. The stomach is the most frequent primary site (~50%), followed by

the small bowel in about a third of cases. As endoscopy for other indications has become more widely used, otherwise asymptomatic and probably clinically insignificant small GISTs have been identified more frequently; it is not clear that the actual incidence has substantially increased. Symptoms associated with GISTs include acute gastrointestinal bleeding leading to melena and/or hematemesis. Anemia may be reflected in generalized weakness. With larger tumors, abdominal distention and pain may be presenting symptoms. At endoscopy, a non-specific smooth bulging mass covered by normal mucosa is the most frequent finding. Initial biopsy may not reveal an epithelial neoplasm. EUS both to assess the extent of the neoplasm and to guide biopsy in order to obtain adequate tissue for histology and molecular pathology may be helpful.

Histologically, a spindle cell neoplasm is the most common subtype (~70%), with epithelioid cells making up 20%; 10% of cases are mixed histology. Immunohistochemical stains for the presence of c-kit or CD34 positivity and mutational analysis of *cKIT* and *PDGFRα* should be performed in all patients. These help to distinguish between GISTs and leiomyoma neoplasms. For nonmetastatic primary GISTs, laparoscopic surgical resection, if feasible, is the treatment of choice; because lymph node metastases are unusual, wedge or segmental resection is preferred. Endoscopic resection has been used in selected cases. Neither histology nor the presence of a *cKIT* or *PDGFRα* mutation distinguishes GISTs with clinically malignant phenotype from those that have a benign course. Higher risk tumors (larger size, higher mitotic index) may present with or develop metastatic or locally unresectable disease. For these patients, use of a c-kit tyrosine kinase inhibitor such as imatinib for tumors with *cKIT* mutations offers effective palliation. Avapritinib is used for tumors with certain *PDGFRα* mutations. However, resistance almost invariably develops, and the development of newer agents effective in tumors with secondary mutations is a high priority. For high-risk GISTs, adjuvant therapy with imatinib for 3 years improves relapse-free and overall survival.

■ SMALL-BOWEL NEOPLASMS

Although the number of new cases of small-bowel neoplasms is substantially less than that of gastric neoplasms (in 2020, there were an estimated 11,110 new cases in the United States, representing 3–4% of gastrointestinal malignancies), the spectrum of malignant tumors of the small bowel is similar and includes NETs (carcinoid), adenocarcinomas, lymphomas, and GISTs. Neuroendocrine cancers are slightly more frequent (40–45%) than adenocarcinomas (30–40%), with the remainder mostly lymphomas and ~8% GISTs. The duodenum is the most common portion of the small bowel in which malignancies develop (~50%), with ~30% occurring in the jejunum and 20% in the ileum. NETs are the most common benign and malignant tumors of the ileum. Risk factors for the development of adenocarcinoma include inflammatory bowel disease (Crohn's disease) and inherited germline mutation syndromes such as Lynch syndrome, familial adenomatous polyposis (FAP), and Peutz-Jeghers syndrome. Celiac disease is associated with a higher incidence of both small-bowel adenocarcinomas and T-cell lymphomas.

While an asymptomatic small-bowel primary adenocarcinoma might be found during surveillance in patients at high risk (such as someone with FAP), symptoms due to obstruction or bleeding (which may be occult) lead to the diagnosis of a small-bowel tumor in a substantial number of patients. Both adenocarcinoma and lymphomas might present with perforation. The development of anemia or obstructive symptoms in a patient with a germline cancer susceptibility gene mutation should lead to a high degree of suspicion for developing malignancy. Evaluation by diagnostic CT imaging may reveal an obstructing lesion. CT and/or MRI enterography can be helpful if the diagnostic CT is not informative. Endoscopy using techniques such as double balloon enteroscopy or video capsule endoscopy allows (for the former) tissue diagnosis as well as localization. Video capsule endoscopy is contraindicated in the setting of obstruction. For NETs, a gallium-68 DOTATATE scan may identify both the primary site as well as metastatic disease. Blood tumor biomarkers are nonspecific for the primary site (CEA or CA19-9 for adenocarcinoma or serum

chromogranin for NET); these assays are better used to monitor response or progression of disease rather than for diagnosis.

Adenocarcinoma of the small bowel appears to be increasing in incidence. The median age for sporadic small-bowel tumors is in the seventh or eighth decade of life, but genetically predisposed patients and those with inflammatory bowel disease may be diagnosed at a much earlier age. African Americans have a higher incidence of small-bowel cancer than whites. While systemic therapies are usually modeled on agents used to treat colorectal cancer, genomic analyses have indicated that small-bowel adenocarcinoma has distinct genomic alterations compared to either colorectal or gastric cancers. Genomic alterations less frequent in small-bowel than in colorectal cancers include *TP53*, *BRAF V600E*, and *APC* mutations, whereas the rate of *KRAS* mutations is similar to colorectal cancer. Within small-bowel sites, the most striking difference is the higher rate of *ERBB2* alterations (of which a minority are amplifications) in duodenal cancers. Not surprisingly, because Lynch syndrome increases the risk of small-bowel adenocarcinoma, 15–20% of these tumors are MSI high or mismatch repair deficient; small-bowel adenocarcinoma associated with celiac disease also may have an increased rate of MSI-high tumors. MSI/MMRP status should be assessed in all patients with small-bowel adenocarcinoma. Somatic tumor genomic analysis may suggest a germline mutation, but appropriate genetic testing for a germline driver mutation should be performed in all patients with small-bowel adenocarcinoma.

Small-bowel adenocarcinoma has its own staging classification in the AJCC eighth edition.

TREATMENT

Small-Bowel Adenocarcinomas

Surgical resection with negative microscopic margins (R0), as is the case for other gastrointestinal tumors, is the best chance for cure. For duodenal adenocarcinomas, a Whipple procedure may be needed; for more distal duodenal cancers and jejunal adenocarcinomas, a segmental resection with adequate margins should be performed. Distal ileal tumors may require right hemicolectomy.

Small-bowel cancers are frequently found with locally advanced disease at the time of diagnosis. If the tumor is resectable, postoperative adjuvant systemic therapy is currently recommended for lymph node-positive patients, using regimens such as capecitabine-oxaliplatin. Benefit from adjuvant therapy has not yet been proven. Small-bowel cancers developing in patients with Lynch syndrome probably have a better prognosis; if colorectal cancer is a model, postoperative adjuvant therapy for patients with Lynch syndrome should be combination chemotherapy, not single-agent fluorinated pyrimidines. The value of immune modulation therapy is not established. For duodenal cancers, chemoradiation is considered if the resection margins are still positive.

For patients with advanced metastatic disease, in the absence of Lynch syndrome or a hypermutated tumor, similar cytotoxic regimens as deployed for gastric or colon cancers have been widely used. For tumors that are MSI high or dMMPR, immunotherapy is indicated; for tumors that are *HER2* amplified or *BRAF* mutated, targeted therapy may be useful.

■ SMALL-BOWEL GASTROINTESTINAL STROMAL TUMORS

Like small-bowel adenocarcinomas, small-bowel GISTs may present with obstruction or bleeding. Diagnostic techniques are those employed for other small-bowel neoplasms. While the pathological criteria for malignant potential are somewhat different than those used for gastric GISTs, postoperative management and treatment of metastatic disease are the same as those described above for gastric GIST.

■ CARCINOID (NEUROENDOCRINE) TUMORS OF THE SMALL BOWEL

Carcinoid tumors are the most common small-bowel neoplasms. For not yet identified reasons, the incidence of small-bowel carcinoid

tumors has markedly increased over the past several decades. Although known risk factors include inherited genetic cancer susceptibility genes such as *MEN1* and neurofibromatosis 1 (*NFI*), these are unlikely to be the cause of the increase (Chap. 84). However, although the incidence has increased, small-bowel carcinoids are still uncommon, with an estimated incidence of approximately nine cases per million in the United States; the disease is more common in African Americans than whites.

Anatomically, the ileum is the most common part of the small bowel affected (~49%), followed by the duodenum and the jejunum. As is the case for gastric carcinoid tumors, grade is based on mitotic rate and/or Ki-67 immunohistochemistry. Histologic differentiation also uses the same criteria as in gastric carcinoid tumors.

In the absence of metastatic disease to the liver in the subgroup of patients whose tumors are functional (i.e., produce a hormone, usually serotonin; a duodenal NET may be a gastrinoma), presenting symptoms may be vague abdominal discomfort until or unless small-bowel obstruction occurs. Carcinoid syndrome may be found in patients whose tumors are diagnosed with already established hepatic metastasis. Because the liver is very efficient at clearing serotonin on first pass, carcinoid syndrome as a result of small-bowel carcinoid tumors usually does not occur in the absence of hepatic metastasis.

Clinical evaluation includes a diagnostic-quality CT or MRI of the abdomen and pelvis. For duodenal carcinoid tumors, upper endoscopy with EUS is also performed, and for jejunal or ileal carcinoid tumors, colonoscopy is performed. Small-bowel imaging may be performed by CT enterography; if an obstructing tumor is suspected, capsule endoscopy should be avoided. The diagnosis of a metastatic NET may be suspected from the radiographic appearance on CT imaging. Somatostatin analog imaging using gallium-68 DOTATATE is helpful in assessing extent of disease in patients whose tumors are somatostatin analog avid, as well as in identifying patients who may benefit from therapy targeting the somatostatin receptor. The AJCC eighth edition cancer staging manual has a specific small-bowel NET TNM stage classification.

The majority of patients present with locoregional disease, with approximately 40% having identified lymph node metastasis. Ten to 15% of patients have metastatic disease (usually to the liver) at the time of initial diagnosis.

Initial management should be surgical resection with curative intent; for patients with extensive adenopathy involving the root of the mesentery, vascular reconstruction may be required. Since small-bowel NETs may involve multiple tumors (15–30% of patients), the entire small bowel should be carefully examined. For patients with functional carcinoid tumors, somatostatin analog therapy using agents such as octreotide should be given before the induction of anesthesia to avoid a carcinoid crisis. For patients with hepatic metastasis, resection or regional therapy including ablation or hepatic artery embolization for functional tumors may provide effective palliation. Carcinoid syndrome may also be palliated by somatostatin receptor targeted therapy in the patients in whom gallium-68 DOTATATE scanning is positive (the majority of patients with carcinoid syndrome), including agents such as octreotide or lanreotide, or by peptide-directed radiation therapy using lutetium-177. Everolimus, an mTOR kinase inhibitor, has modest activity in metastatic small-bowel carcinoid tumors.

BENIGN NEOPLASMS OF THE SMALL BOWEL

As is the case for malignant small-bowel tumors, benign neoplasms of the small bowel are rare. In addition to cancers, the precursor lesion adenomas or hamartomas (from which cancers develop) may be driven by inherited cancer susceptibility genes (FAP, Lynch syndrome, Peutz-Jeghers syndrome, among others.). Other benign neoplasms include lipomas, leiomyomas, neurofibromas, and benign lymphoid nodular hyperplasia.

Patients with benign small-bowel neoplasms not associated with an inherited cancer susceptibility gene (in which case a benign tumor may be found during surveillance) are usually asymptomatic. A mass may be noted on an imaging study (usually a CT) ordered for another reason. Workup for occult or overt bleeding or intussusception of the bowel may lead to the discovery of a benign small-bowel neoplasm.

Diagnostic evaluation may include video capsule endoscopy and double balloon or push enteroscopy. In general, benign neoplasms, if found during surveillance, are removed endoscopically, if technically feasible, to decrease the risk of intussusception in Peutz-Jeghers syndrome. Mucosectomy may be used to treat bleeding hemangiomas.

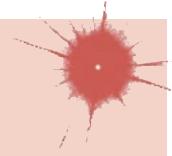
FURTHER READING

- CANCER GENOME ATLAS RESEARCH NETWORK: Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513:202, 2014.
- CANCER GENOME ATLAS RESEARCH NETWORK et al: Integrated genomic characterization of oesophageal carcinoma. *Nature* 541:169, 2017.
- CHOI IJ et al: *Helicobacter pylori* and prevention of gastric cancer. *N Engl J Med* 378:2244, 2018.
- CHOI IJ et al: Family history of gastric cancer and *Helicobacter pylori* treatment. *N Engl J Med* 382:427, 2020.
- CRISTESCU R et al: Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* 21:449, 2015.
- CUNNINGHAM D et al: Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med* 355:11, 2006.
- GOETZE OT et al: Multimodal treatment in locally advanced gastric cancer. *Updates Surg* 70:173, 2018.
- POURMAND K, ITZKOWITZ SH: Small bowel neoplasms and polyps. *Curr Gastroenterol Rep* 18:23, 2016.
- WATANABE M et al: Recent progress in multidisciplinary treatment for patients with esophageal cancer. *Surg Today* 50:12, 2020.

81

Lower Gastrointestinal Cancers

Robert J. Mayer



Lower gastrointestinal cancers include malignant tumors of the colon, rectum, and anus.

COLORECTAL CANCER

INCIDENCE

Cancer of the large bowel is second only to lung cancer as a cause of cancer death in the United States: 149,500 new cases were identified in 2021, and 52,980 deaths were due to colorectal cancer. The incidence rate has decreased significantly during the past 30 years in individuals 50 years of age or older, likely due in large part to enhanced and more compliantly followed screening practices. Similarly, mortality rates in the United States in this age group have decreased by more than 30%, resulting largely from early detection and improved treatment. During the same period of time, however, the incidence rate for colorectal cancer in men and women <50 years of age, having no enhanced genetic risk factor or family history for the disease, has risen by 2% each year with the presence of symptoms in this age group often being initially attributed to other causes, resulting in a more advanced disease stage at the time of diagnosis being frequently observed. A corresponding increase in mortality rates from colorectal cancer in this young adult population is now evident while, simultaneously, the trend for a decreased death rate in older individuals has continued. No distinguishing etiologic or molecular factor or clinical characteristic to account for the rising incidence of colorectal cancer in younger men and women has yet been identified with lifestyle patterns, such as diet and obesity beginning at an earlier age, having been proposed.

POLYPS AND MOLECULAR PATHOGENESIS

Most colorectal cancers, regardless of etiology, arise from adenomatous polyps. A polyp is a grossly visible protrusion from the mucosal surface and may be classified pathologically as a nonneoplastic hamartoma (e.g., *juvenile polyp*), a hyperplastic mucosal proliferation (*hyperplastic polyp*), or an adenomatous polyp. Only adenomas are clearly premalignant, and only a minority of adenomatous polyps evolve into cancer. Adenomatous polyps may be found in the colons of ~30% of middle-aged and ~50% of elderly people; however, <1% of polyps ever become malignant. Most polyps produce no symptoms and remain clinically undetected. Occult blood in the stool is found in <5% of patients with polyps.

A number of molecular changes are noted in adenomatous polyps and colorectal cancers that are thought to reflect a multistep process in the evolution of normal colonic mucosa to life-threatening invasive carcinoma. These developmental steps toward carcinogenesis include, but are not restricted to, point mutations in the *K-ras* proto-oncogene; hypomethylation of DNA, leading to gene activation; loss of DNA (*allelic loss*) at the site of a tumor-suppressor gene (the adenomatous polyposis coli [*APC*] gene) on the long arm of chromosome 5 (5q21); allelic loss at the site of a tumor-suppressor gene located on chromosome 18q (the deleted in colorectal cancer [*DCC*] gene); and allelic loss at chromosome 17p, associated with mutations in the *p53* tumor-suppressor gene (see Fig. 71-2). Thus, the altered proliferative pattern of the colonic mucosa, which results in progression to a polyp and then to carcinoma, may involve the mutational activation of an oncogene followed by and coupled with the loss of genes that normally suppress tumorigenesis. It remains uncertain whether the genetic aberrations always occur in a defined order. Based on this model, however, cancer is believed to develop only in those polyps in which most (if not all) of these mutational events take place.

Clinically, the probability of an adenomatous polyp becoming a cancer depends on the gross appearance of the lesion, its histologic features, and its size. Polyps may be pedunculated (stalked) or sessile (flat-based), adenomatous or serrated. Invasive cancers develop more frequently in sessile, serrated (i.e., "flat") polyps. Histologically, adenomatous polyps may be tubular, villous (i.e., papillary), or tubulovillous. Villous adenomas, most of which are sessile, become malignant more than three times as often as tubular adenomas. The likelihood that any polypoid lesion in the large bowel contains invasive cancer is related to the size of the polyp, being negligible (<2%) in lesions <1.5 cm, intermediate (2–10%) in lesions 1.5–2.5 cm, and substantial (10%) in lesions >2.5 cm in size.

Following the detection of an adenomatous polyp, the entire large bowel should be visualized endoscopically because synchronous lesions are noted in about one-third of cases. Colonoscopy should then be repeated periodically, even in the absence of a previously documented malignancy, because such patients have a 30–50% probability of developing another adenoma and are at a higher-than-average risk for developing a colorectal carcinoma. Adenomatous polyps are thought to require >5 years of growth before becoming clinically significant; colonoscopy need not be carried out more frequently than every 3 years for the vast majority of patients.

ETIOLOGY AND RISK FACTORS

Risk factors for the development of colorectal cancer are listed in **Table 81-1**.

TABLE 81-1 Risk Factors for the Development of Colorectal Cancer

Diet: Animal fat, obesity
Hereditary syndromes
Polyposis coli
MYH-associated polyposis
Nonpolyposis syndrome (Lynch's syndrome)
Inflammatory bowel disease
<i>Streptococcus bovis</i> bacteremia
Tobacco use

Diet The etiology for most cases of large-bowel cancer appears to be related to environmental factors. The disease occurs more often in upper socioeconomic populations who live in urban areas. Mortality from colorectal cancer is directly correlated with per capita consumption of calories, meat protein, and dietary fat and oil as well as elevations in the serum cholesterol concentration and mortality from coronary artery disease. Geographic variations in incidence largely are unrelated to genetic differences because migrant groups tend to assume the large-bowel cancer incidence rates of their adopted countries. Furthermore, population groups such as Mormons and Seventh Day Adventists, whose lifestyle and dietary habits differ somewhat from those of their neighbors, have significantly lower-than-expected incidence and mortality rates for colorectal cancer. The incidence of colorectal cancer has increased in Japan since that nation has adopted a more "Western" diet. At least three hypotheses have been proposed to explain the relationship to diet, none of which is fully satisfactory.

ANIMAL FATS One hypothesis is that the ingestion of animal fats found in red meats and processed meat leads to an increased proportion of anaerobes in the gut microflora (the "microbiome"), resulting in the conversion of normal bile acids into carcinogens. This provocative hypothesis is supported by several reports of increased amounts of fecal anaerobes (*Fusobacterium nucleatum*, *Bacteroides fragilis*) in the stools of patients with colorectal cancer. Diets high in animal (but not vegetable) fats are also associated with high serum cholesterol, which is also associated with enhanced risk for the development of colorectal adenomas and carcinomas.

INSULIN RESISTANCE The large number of calories in Western diets coupled with physical inactivity has been associated with a higher prevalence of obesity. Obese persons develop insulin resistance with increased circulating levels of insulin, leading to higher circulating concentrations of insulin-like growth factor type I (IGF-I). This growth factor appears to stimulate proliferation of the intestinal mucosa.

FIBER Contrary to prior beliefs, the results of randomized trials and case-controlled studies have failed to show any value for dietary fiber or diets high in fruits and vegetables in preventing the recurrence of colorectal adenomas or the development of colorectal cancer.

The weight of epidemiologic evidence, however, implicates diet as being the major etiologic factor for colorectal cancer, particularly diets high in animal fat and in calories.

HEREDITARY FACTORS AND SYNDROMES

Up to 25% of patients with colorectal cancer have a family history of the disease, suggesting a hereditary predisposition. Inherited large-bowel cancers can be divided into two main groups: the well-studied but uncommon polyposis syndromes and the more common nonpolyposis syndromes (**Table 81-2**).

Polyposis Coli Polyposis coli (familial polyposis of the colon) is a rare condition characterized by the appearance of thousands of adenomatous polyps throughout the large bowel. It is transmitted as an autosomal dominant trait; the occasional patient with no family history probably developed the condition due to a spontaneous mutation. Polyposis coli is associated with a deletion in the long arm of chromosome 5 (including the *APC* gene) in both neoplastic (somatic mutation) and normal (germline mutation) cells. The loss of this genetic material (i.e., allelic loss) results in the absence of tumor-suppressor genes whose protein products would normally inhibit neoplastic growth. The presence of soft tissue and bony tumors, congenital hypertrophy of the retinal pigment epithelium, mesenteric desmoid tumors, and ampullary cancers in addition to the colonic polyps characterizes a subset of polyposis coli known as *Gardner's syndrome*. The appearance of malignant tumors of the central nervous system accompanying polyposis coli defines *Turcot's syndrome*. The colonic polyps in all these conditions are rarely present before puberty but are generally evident in affected individuals by age 25 years. If the polyposis is not treated surgically, colorectal cancer will develop in almost all patients aged <40 years. Polyposis coli results from a defect in the colonic mucosa, leading to an abnormal proliferative pattern and impaired DNA repair mechanisms.

TABLE 81-2 Heritable (Autosomal Dominant) Gastrointestinal Neoplasia Syndromes

SYNDROME	DISTRIBUTION OF POLYPS	HISTOLOGIC TYPE	MALIGNANT POTENTIAL	ASSOCIATED LESIONS
Familial adenomatous polyposis	Large intestine	Adenoma	Common	None
Gardner's syndrome	Large and small intestines	Adenoma	Common	Osteomas, fibromas, lipomas, epidermoid cysts, ampullary cancers, congenital hypertrophy of retinal pigment epithelium
Turcot's syndrome	Large intestine	Adenoma	Common	Brain tumors
MYH-associated polyposis	Large intestine	Adenoma	Common	None
Lynch syndrome (nonpolyposis syndrome)	Large intestine (often proximal)	Adenoma	Common	Endometrial and ovarian tumors (most frequently), gastric, genitourinary, pancreatic, biliary cancers (less frequently)
Peutz-Jeghers syndrome	Small and large intestines, stomach	Hamartoma	Rare	Mucocutaneous pigmentation; tumors of the ovary, breast, pancreas, endometrium
Juvenile polyposis	Large and small intestines, stomach	Hamartoma, rarely progressing to adenoma	Rare	Various congenital abnormalities

Once the multiple polyps are detected, patients should undergo a total colectomy. Medical therapy with nonsteroidal anti-inflammatory drugs (NSAIDs) such as sulindac and selective cyclooxygenase-2 inhibitors such as celecoxib can decrease the number and size of polyps in patients with polyposis coli; however, this effect on polyps is only temporary, and the use of NSAIDs has not been shown to reduce the risk of cancer. Colectomy remains the primary therapy/prevention. The offspring of patients with polyposis coli, who often are prepubertal when the diagnosis is made in the parent, have a 50% risk for developing this premalignant disorder and should be carefully screened by annual flexible sigmoidoscopy until age 35 years. Proctosigmoidoscopy is a sufficient screening procedure because polyps tend to be evenly distributed from cecum to anus, making more invasive and expensive techniques such as colonoscopy or barium enema unnecessary. Testing for occult blood in the stool is an inadequate screening maneuver. If a causative germline APC mutation has been identified in an affected family member, an alternative method for identifying carriers is testing DNA from peripheral blood mononuclear cells for the presence of the specific APC mutation. The detection of such a germline mutation can lead to a definitive diagnosis before the development of polyps.

MYH-Associated Polyposis MYH-associated polyposis (MAP) is a rare autosomal recessive syndrome caused by a biallelic mutation in the *MUT4H* gene. This hereditary condition may have a variable clinical presentation, resembling polyposis coli or colorectal cancer occurring in younger individuals without polyposis. Screening and colectomy guidelines for this syndrome are less clear than for polyposis coli, but annual to biennial colonoscopic surveillance is generally recommended starting at age 25–30 years.

Lynch Syndrome Lynch syndrome, previously known as hereditary nonpolyposis colon cancer, is another autosomal dominant trait. It is characterized by the presence of three or more relatives with histologically documented colorectal cancer, one of whom is a first-degree relative of the other two; one or more cases of colorectal cancer diagnosed before age 50 in the family; and colorectal cancer involving at least two generations. In contrast to polyposis coli, Lynch syndrome is associated with an unusually high frequency of cancer arising in the proximal large bowel. The median age for the appearance of an adenocarcinoma is <50 years, 10–15 years younger than the median age for the general population. Despite having a poorly differentiated, mucinous histologic appearance, the proximal colon tumors that characterize Lynch syndrome have a better prognosis than sporadic tumors from patients of similar age. Families with Lynch syndrome often include individuals with multiple primary cancers; the association of colorectal cancer with either ovarian or endometrial carcinomas is especially strong in women, and an increased appearance of gastric, small-bowel, genitourinary, pancreaticobiliary, and sebaceous skin tumors has been

reported as well. It has been recommended that members of such families undergo annual or biennial colonoscopy beginning at age 25 years, with intermittent pelvic ultrasonography and endometrial biopsy for afflicted women; such a screening strategy has not yet been validated. Lynch syndrome is associated with germline mutations of several genes, particularly *hMSH2* on chromosome 2 and *hMLH1* on chromosome 3. These mutations lead to errors in DNA replication and are thought to result in DNA instability because of defective repair of DNA mismatches resulting in abnormal cell growth and tumor development. Testing tumor cells through molecular analysis of DNA for “microsatellite instability” or immunohistochemical staining for deficiency in mismatch repair proteins in patients with colorectal cancer and a positive family history for colorectal or endometrial cancer may identify probands with Lynch syndrome.

■ INFLAMMATORY BOWEL DISEASE

(Chap. 326) Large-bowel cancer is increased in incidence in patients with long-standing inflammatory bowel disease (IBD). Cancers develop more commonly in patients with ulcerative colitis than in those with granulomatous (i.e., Crohn's) colitis, but this impression may result in part from the occasional difficulty of differentiating these two conditions. The risk of colorectal cancer in a patient with IBD is relatively small during the initial 10 years of the disease, but then appears to increase at a rate of ~0.5–1% per year. Cancer may develop in 8–30% of patients after 25 years. The risk is higher in younger patients with pancolitis.

Cancer surveillance strategies in patients with IBD are unsatisfactory. Symptoms such as bloody diarrhea, abdominal cramping, and obstruction, which may signal the appearance of a tumor, are similar to the complaints caused by a flare-up of the underlying inflammatory disease. In patients with a history of IBD lasting ≥15 years who continue to experience exacerbations, the surgical removal of the colon can significantly reduce the risk for cancer and also eliminate the target organ for the underlying chronic gastrointestinal disorder. The value of such surveillance techniques as colonoscopy with mucosal biopsies and brushings for less symptomatic individuals with chronic IBD is uncertain. The lack of uniformity regarding the pathologic criteria that characterize dysplasia and the absence of data that such surveillance reduces the development of lethal cancers have made this costly practice an area of controversy.

■ OTHER HIGH-RISK CONDITIONS

***Streptococcus bovis* Bacteremia** For unknown reasons, individuals who develop endocarditis or septicemia from this fecal bacterium have a high incidence of occult colorectal tumors and, possibly, upper gastrointestinal cancers as well. Endoscopic or radiographic screening appears advisable.

Tobacco Use Cigarette smoking is linked to the development of colorectal adenomas, particularly after >35 years of tobacco use. No biologic explanation for this association has yet been proposed.

■ PRIMARY PREVENTION

Several orally administered compounds have been assessed as possible inhibitors of colon cancer. The most effective class of chemopreventive agents is aspirin and other NSAIDs, which are thought to suppress cell proliferation by inhibiting prostaglandin synthesis. Regular aspirin use reduces the risk of colon adenomas and carcinomas as well as death from large-bowel cancer; such use also appears to diminish the likelihood for developing additional premalignant adenomas following successful treatment for a prior colon carcinoma. This effect of aspirin on colon carcinogenesis increases with the duration and dosage of drug use. Emerging data linking adequate plasma levels of vitamin D with reduced risk of adenomatous polyps and colorectal cancer appear promising. The value of vitamin D as a form of chemoprevention is under study. Antioxidant vitamins such as ascorbic acid, tocopherols, and β-carotene are ineffective at reducing the incidence of subsequent adenomas in patients who have undergone the removal of a colon adenoma. Estrogen replacement therapy has been associated with a reduction in the risk of colorectal cancer in women, conceivably by an effect on bile acid synthesis and composition or by decreasing synthesis of IGF-I.

■ SCREENING

The rationale for colorectal cancer screening programs is that the removal of adenomatous polyps will prevent colorectal cancer, and that earlier detection of localized, superficial cancers in asymptomatic individuals will increase the surgical cure rate. Such screening programs are particularly important for individuals with a family history of the disease in first-degree relatives. The relative risk for developing colorectal cancer increases to 1.75 in such individuals and may be even higher if the relative was afflicted before age 60 years. The prior use of rigid proctosigmoidoscopy as a screening tool was based on the observation that 60% of early lesions are located in the rectosigmoid. For unexplained reasons, however, the proportion of large-bowel cancers arising in the rectum has been decreasing during the past several decades, with a corresponding increase in the proportion of cancers in the more proximal descending colon. As such, the potential for rigid proctosigmoidoscopy to detect a sufficient number of occult neoplasms to make the procedure cost-effective has been questioned.

Screening strategies for colorectal cancer that have been examined during the past several decades are listed in **Table 81-3**.

Many programs directed at the early detection of colorectal cancers have focused on digital rectal examinations and fecal occult blood (i.e., stool guaiac) testing. The digital examination should be part of any routine physical evaluation in adults aged >40 years, serving as a screening test for prostate cancer in men, a component of the pelvic examination in women, and an inexpensive maneuver for the detection of masses in the rectum. However, because of the proximal migration of colorectal tumors, its value as an overall screening modality for colorectal cancer has become limited. The development of the fecal occult blood test has greatly facilitated the detection of occult fecal blood. Unfortunately, even when performed optimally, the fecal occult blood test has major

limitations as a screening technique. About 50% of patients with documented colorectal cancers have a negative fecal occult blood test, consistent with the intermittent bleeding pattern of these tumors. When random cohorts of asymptomatic persons have been tested, 2–4% have fecal occult blood-positive stools. Colorectal cancers have been found in <10% of these “test-positive” cases, with benign polyps being detected in an additional 20–30%. Thus, a colorectal neoplasm will not be found in most asymptomatic individuals with occult blood in their stool. Nonetheless, persons found to have fecal occult blood-positive stool routinely undergo further medical evaluation, including sigmoidoscopy and/or colonoscopy—procedures that are not only uncomfortable and expensive but also associated with a small risk for significant complications. The added cost of these studies would appear justifiable if the small number of patients found to have occult neoplasms because of fecal occult blood screening could be shown to have an improved prognosis and prolonged survival. Prospectively controlled trials have shown a statistically significant reduction in mortality rate from colorectal cancer for individuals undergoing annual stool guaiac screening. However, this benefit only emerged after >13 years of follow-up and was extremely expensive to achieve because all positive tests (most of which were falsely positive) were followed by colonoscopy. Moreover, these colonoscopic examinations quite likely provided the opportunity for cancer prevention through the removal of potentially premalignant adenomatous polyps because the eventual development of cancer was reduced by 20% in the cohort undergoing annual screening.

With the appreciation that the carcinogenic process leading to the progression of the normal bowel mucosa to an adenomatous polyp and then to a cancer is the result of a series of molecular changes, investigators have examined fecal DNA for evidence of mutations associated with such molecular changes as evidence of the occult presence of precancerous lesions or actual malignancies. Such a strategy has been tested in >4000 asymptomatic individuals whose stool was assessed for occult blood and for 21 possible mutations in fecal DNA; these study subjects also underwent colonoscopy. Although the fecal DNA strategy proved to be more effective for suggesting the presence of more advanced adenomas and cancers than did the fecal occult blood testing approach, the overall sensitivity, using colonoscopic findings as the standard, was <50%, diminishing enthusiasm for further pursuit of the fecal DNA screening strategy.

The use of imaging studies to screen for colorectal cancers has also been explored. Air contrast barium enemas had been used to identify sources of occult blood in the stool prior to the advent of fiberoptic endoscopy; the cumbersome nature of the procedure and inconvenience to patients limited its widespread adoption. The introduction of CT scanning led to the development of virtual (i.e., CT) colonography as an alternative to the growing use of endoscopic screening techniques. Virtual colonography was proposed as being equivalent in sensitivity to colonoscopy and being available in a more widespread manner because it did not require the same degree of operator expertise as fiberoptic endoscopy. However, virtual colonography requires the same cathartic preparation that has limited widespread acceptance in association with endoscopic colonoscopy, is diagnostic but not therapeutic (i.e., patients with suspicious findings must undergo a subsequent endoscopic procedure for polypectomy or biopsy), and, in the setting of general radiology practices, appears to be less sensitive as a screening technique when compared with endoscopic procedures.

With the appreciation of the inadequacy of fecal occult blood testing alone, concerns about the practicality of imaging approaches, and the wider adoption of endoscopic examinations by the primary care community, screening strategies in asymptomatic persons have changed. At present, the American Cancer Society, the American College of Gastroenterology, and the National Comprehensive Cancer Network recommend either fecal occult blood testing annually coupled with flexible sigmoidoscopy every 5 years or colonoscopy every 10 years in asymptomatic individuals with no personal or family history of polyps or colorectal cancer. In view of the emerging increase in the incidence of colorectal cancer in individuals <50 years of age, guidelines issued from these organizations have recently lowered the age at which to begin such screening from age 50 to age 45 years. The recommendation

TABLE 81-3 Screening Strategies for Colorectal Cancer

Digital rectal examination
Stool testing
• Occult blood
• Fecal DNA
Imaging
• Contrast barium enema
• Virtual (i.e., computed tomography colonography)
Endoscopy
• Flexible sigmoidoscopy
• Colonoscopy

for the inclusion of flexible sigmoidoscopy is strongly supported by the recently published results of three randomized trials performed in the United States, the United Kingdom, and Italy, involving >350,000 individuals, which consistently showed that periodic (even single) sigmoidoscopic examinations, after more than a decade of median follow-up, lead to an ~21% reduction in the development of colorectal cancer and a >25% reduction in mortality from the malignant disease. Less than 20% of participants in these studies underwent a subsequent colonoscopy. In contrast to the cathartic preparation required before colonoscopic procedures, which is only performed by highly trained specialists, flexible sigmoidoscopy requires only an enema as preparation and can be accurately performed by nonspecialty physicians or physician-extenders. The randomized screening studies using flexible sigmoidoscopy led to the estimate that ~650 individuals needed to be screened to prevent one colorectal cancer death; this contrasts with the data for mammography where the number of women needing to be screened to prevent one breast cancer death is 2500, reinforcing the efficacy of endoscopic surveillance for colorectal cancer screening. Presumably the benefit from the sigmoidoscopic screening is the result of the identification and removal of adenomatous polyps; it is intriguing that this benefit has been achieved using a technique that leaves the proximal half of the large bowel unvisualized.

It remains to be seen whether surveillance colonoscopy, which has gained increasing popularity in the United States for colorectal cancer screening, will prove to be more effective than flexible sigmoidoscopy in reducing mortality from this disease. Ongoing randomized trials being conducted in Europe are addressing this issue. Although flexible sigmoidoscopy only visualizes the distal half of the large bowel, leading to the assumption that colonoscopy represents a more informative approach, colonoscopy has been reported as being less accurate for screening the proximal rather than the distal colon, perhaps due to technical considerations but also possibly because of a greater frequency of serrated (i.e., "flat") polyps in the right colon, which are more difficult to identify. Furthermore, the vast majority of colorectal cancers that have appeared in younger adults have arisen in the left colon (i.e., distal to the splenic flexure), within the visible range of a flexible sigmoidoscopy. At present, colonoscopy performed every 10 years has been offered as an alternative to annual fecal occult blood testing with periodic (every 5 years) flexible sigmoidoscopy. Colonoscopy has been shown to be superior to double-contrast barium enema and also to have a higher sensitivity for detecting villous or dysplastic adenomas or cancers than the strategy using occult fecal blood testing and flexible sigmoidoscopy. Whether colonoscopy performed every 10 years beginning at age 45 is medically superior and economically equivalent to flexible sigmoidoscopy remains to be determined.

CLINICAL FEATURES

Presenting Symptoms Symptoms vary with the anatomic location of the tumor. Because stool is relatively liquid as it passes through the ileocecal valve into the right colon, cancers arising in the cecum and ascending colon may become quite large without resulting in any obstructive symptoms or noticeable alterations in bowel habits. Lesions of the right colon commonly ulcerate, leading to chronic, insidious blood loss without a change in the appearance of the stool. Consequently, patients with tumors of the ascending colon often present with symptoms such as fatigue, palpitations, and even angina pectoris and are found to have a hypochromic, microcytic anemia, indicative of iron deficiency. Because the cancers may bleed intermittently, a random fecal occult blood test may be negative. As a result, the unexplained presence of iron-deficiency anemia in any adult (with the possible exception of a premenopausal, multiparous woman) mandates a thorough endoscopic and/or radiographic visualization of the entire large bowel (**Fig. 81-1**).

Because stool becomes more formed as it passes into the transverse and descending colon, tumors arising there tend to impede the passage of stool, resulting in the development of abdominal cramping, occasional obstruction, and even perforation. Radiographs of the abdomen often reveal characteristic annular, constricting lesions ("apple-core" or "napkin-ring") (**Fig. 81-2**).



FIGURE 81-1 Double-contrast air-barium enema revealing a sessile tumor of the cecum in a patient with iron-deficiency anemia and guaiac-positive stool. The lesion at surgery was a stage II adenocarcinoma.

Cancers arising in the rectosigmoid are often associated with hematochezia, tenesmus, and narrowing of the caliber of stool; anemia is an infrequent finding. While these symptoms may lead patients and their physicians to suspect the presence of hemorrhoids, the development of rectal bleeding and/or altered bowel habits demands a prompt digital rectal examination and proctosigmoidoscopy.

Staging, Prognostic Factors, and Patterns of Spread The prognosis for individuals having colorectal cancer is related to the depth of tumor penetration into the bowel wall and the presence of both regional lymph node involvement and distant metastases. These variables are incorporated into a TNM classification method, in which T represents the depth of tumor penetration, N the presence of lymph node involvement, and M the presence or absence of distant metastases



FIGURE 81-2 Annular, constricting adenocarcinoma of the descending colon. This radiographic appearance is referred to as an "apple-core" lesion and is always highly suggestive of malignancy.

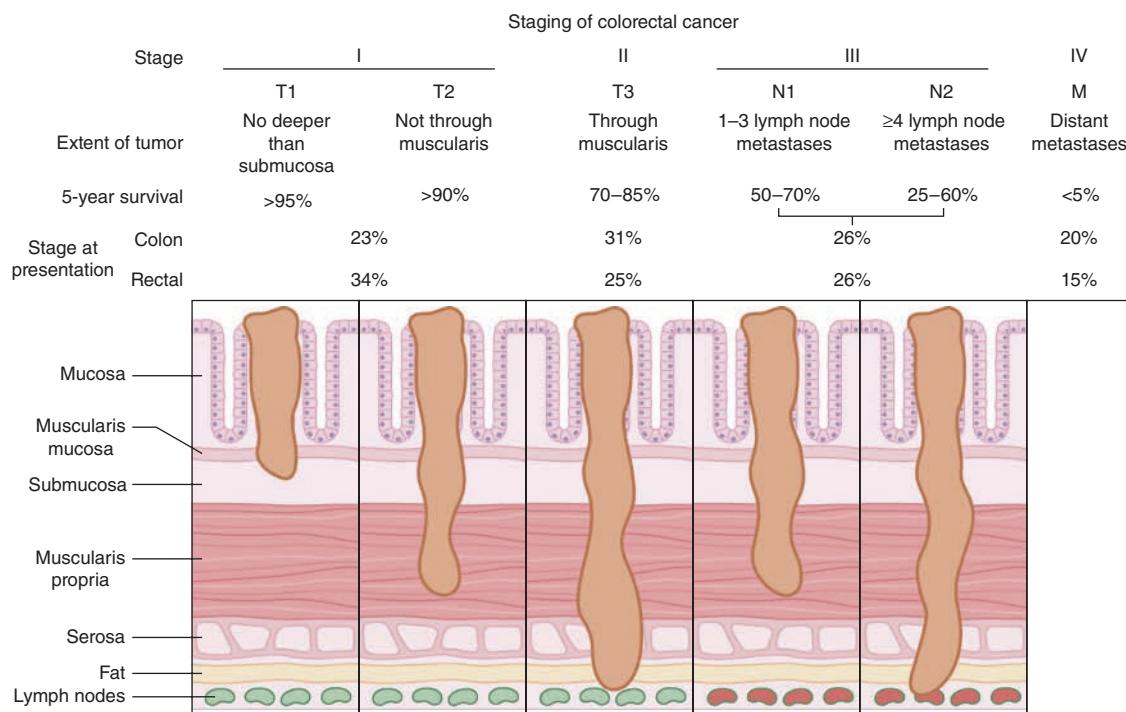


FIGURE 81-3 Staging and prognosis for patients with colorectal cancer.

(Fig. 81-3). Superficial lesions that do not involve regional lymph nodes and do not penetrate through the submucosa (T1) or the muscularis (T2) are designated as *stage I* (T1–2N0M0) disease; tumors that penetrate through the muscularis but have not spread to lymph nodes are *stage II* disease (T3–4N0M0); regional lymph node involvement defines *stage III* (TXN1–2M0) disease; and metastatic spread to sites such as liver, lung, or bone indicates *stage IV* (TXNM1) disease. Unless gross evidence of metastatic disease is present, disease stage cannot be determined accurately before surgical resection and pathologic analysis of the operative specimens.

Most recurrences after a surgical resection of a large-bowel cancer occur within the first 4 years, making 5-year survival a fairly reliable indicator of cure. The likelihood for 5-year survival in patients with colorectal cancer is stage-related (Fig. 81-3). That likelihood has improved during the past several decades when similar surgical stages have been compared. The most plausible explanation for this improvement is more thorough intraoperative and pathologic staging. In particular, more exacting attention to pathologic detail has revealed that the prognosis following the resection of a colorectal cancer is not related merely to the presence or absence of regional lymph node involvement; rather, prognosis may be more precisely gauged by the number of involved lymph nodes (one to three lymph nodes [“N1”] vs four or more lymph nodes [“N2”]) and the number of nodes examined. A minimum of 12 sampled lymph nodes is thought necessary to accurately define tumor stage, and the more nodes examined, the better. Other predictors of a poor prognosis after a total surgical resection include tumor penetration through the bowel wall into pericolic fat, poorly differentiated histology, perforation and/or tumor adherence to adjacent organs (increasing the risk for an anatomically adjacent recurrence), and venous invasion by tumor (Table 81-4). Regardless of the clinicopathologic stage, a preoperative elevation of the plasma carcinoembryonic antigen (CEA) level predicts eventual tumor recurrence. The presence of specific chromosomal aberrations, particularly a mutation in the *b-raf* gene in tumor cells, appears to predict for a higher risk for metastatic spread. Conversely, the detection of microsatellite instability in tumor tissue indicates a more favorable outcome. Tumors arising in the left colon are associated with a better prognosis than those appearing in the right colon, likely due to differences in molecular patterns. In contrast to most other cancers, the prognosis in colorectal cancer is not influenced by the size of the primary lesion when adjusted for nodal involvement and histologic differentiation.

Cancers of the large bowel generally spread to regional lymph nodes or to the liver via the portal venous circulation. The liver represents the most frequent visceral site of metastasis; it is the initial site of distant spread in one-third of recurring colorectal cancers and is involved in more than two-thirds of such patients at the time of death. In general, colorectal cancer rarely spreads to the lungs, supraclavicular lymph nodes, bone, or brain without prior spread to the liver. A major exception to this rule occurs in patients having primary tumors in the distal rectum, from which tumor cells may spread through the paravertebral venous plexus, escaping the portal venous system and thereby reaching the lungs or supraclavicular lymph nodes without hepatic involvement. The median survival after the detection of distant metastases has increased during the last 30 years from 6–9 months (hepatomegaly, abnormal liver chemistries) to 27–30 months (small liver nodule initially identified by elevated CEA level and subsequent CT scan) with increasingly effective systemic therapy improving this prognosis further.

Efforts to use gene expression profiles to identify patients at risk of recurrence or those particularly likely to benefit from adjuvant therapy have not yet yielded practice-changing results. Despite a burgeoning literature examining a host of prognostic factors, pathologic stage at diagnosis remains the best predictor of long-term prognosis. Patients with lymphovascular invasion and high preoperative CEA levels are likely to have a more aggressive clinical course.

TABLE 81-4 Predictors of Poorer Outcomes Following Total Surgical Resection of Colorectal Cancer

Tumor spread to regional lymph nodes
Number of regional lymph nodes involved
Tumor penetration through the bowel wall
Poorly differentiated histology
Perforation
Tumor adherence to adjacent organs
Venous invasion
Preoperative elevation of CEA titer (>5 ng/mL)
Specific chromosomal deletion (e.g., mutation in the <i>b-raf</i> gene)
Right-sided location of primary tumor

Abbreviation: CEA, carcinoembryonic antigen.

TREATMENT

Colorectal Cancer

Total resection of tumor is the optimal treatment when a malignant lesion is detected in the large bowel. An evaluation for the presence of metastatic disease, including a thorough physical examination, biochemical assessment of liver function, measurement of the plasma CEA level, and a CT scan of the chest, abdomen, and pelvis, should be performed before surgery. When possible, a colonoscopy of the entire large bowel should be performed to identify synchronous neoplasms and/or polyps. The detection of metastases should not preclude surgery in patients with tumor-related symptoms such as gastrointestinal bleeding or obstruction, but it often prompts the use of a less radical operative procedure. The necessity for a primary tumor resection in asymptomatic individuals with metastatic disease is an area of controversy. At the time of laparotomy, the entire peritoneal cavity should be examined, with thorough inspection of the liver, pelvis, and hemidiaphragm and careful palpation of the full length of the large bowel. Following recovery from a complete resection, patients should be observed carefully for 5 years by semi-annual physical examinations and blood chemistry measurements. If a complete colonoscopy was not performed preoperatively, it should be carried out within the first several postoperative months. Some authorities favor measuring plasma CEA levels at 3-month intervals because of the sensitivity of this test as a marker for otherwise undetectable tumor recurrence. The value of periodically assessing plasma for the presence of circulating tumor DNA as a biomarker for residual or recurrent disease is under study. Subsequent endoscopic surveillance of the large bowel, probably at triennial intervals, is indicated, because patients who have been cured of one colorectal cancer have a 3–5% probability of developing an additional bowel cancer during their lifetime and a >15% risk for the development of adenomatous polyps. Anastomotic (“suture-line”) recurrences are infrequent in colorectal cancer patients, provided the surgical resection margins were adequate and free of tumor. The value of periodic CT scans of the abdomen, assessing for an early, asymptomatic indication of tumor recurrence, while uncertain, has been recommended annually for the first 3 postoperative years.

Radiation therapy to the pelvis is recommended for patients with rectal cancer because it reduces the 20–25% probability of regional recurrences following complete surgical resection of stage II or III tumors, especially if they have penetrated through the serosa. This alarmingly high rate of local disease recurrence is believed to be due to the fact that the contained anatomic space within the pelvis limits the extent of the resection and because the rich lymphatic network of the pelvic side wall immediately adjacent to the rectum facilitates the early spread of malignant cells into surgically inaccessible tissue. The use of sharp rather than blunt dissection of rectal cancers (*total mesorectal excision*) appears to reduce the likelihood of local disease recurrence to ~10%. Radiation therapy, either administered pre- or postoperatively, further reduces the likelihood of pelvic recurrences but does not appear to prolong survival. Combining radiation therapy with 5-fluorouracil (5-FU)-based chemotherapy, preferably prior to surgical resection, lowers local recurrence rates and improves overall survival. Radiation therapy alone is not effective as the primary treatment of colon cancer.

Systemic therapy for patients with colorectal cancer has become more effective. 5-FU remains the backbone of treatment for this disease. Partial responses are obtained in 15–20% of patients. The probability of tumor response appears to be somewhat greater for patients with liver metastases when chemotherapy is infused directly into the hepatic artery, but intraarterial treatment is costly and toxic and does not appear to appreciably prolong survival. The concomitant administration of folinic acid (leucovorin [LV]) improves the efficacy of 5-FU in patients with advanced colorectal cancer, presumably by enhancing the binding of 5-FU to its target enzyme, thymidylate synthase. 5-FU is generally administered

intravenously but may also be given orally in the form of capecitabine (Xeloda) with seemingly similar efficacy.

Irinotecan (CPT-11), a topoisomerase 1 inhibitor, has been added to 5-FU and LV (e.g., FOLFIRI) with resultant improvement in response rates and survival of patients with metastatic disease. The *FOLFIRI regimen* is as follows: irinotecan, 180 mg/m² as a 90-min infusion on day 1; LV, 400 mg/m² as a 2-h infusion during irinotecan administration; immediately followed by 5-FU bolus, 400 mg/m², and 46-h continuous infusion of 2.4–3 g/m² every 2 weeks. Diarrhea is the major side effect from irinotecan. Oxaliplatin, a platinum analogue, also improves the response rate when added to 5-FU and LV (FOLFOX) as initial treatment of patients with metastatic disease. The *FOLFOX regimen* is as follows: 2-h infusion of LV (400 mg/m² per day) followed by a 5-FU bolus (400 mg/m² per day) and 22-h infusion (1200 mg/m²) every 2 weeks, together with oxaliplatin, 85 mg/m² as a 2-h infusion on day 1. Oxaliplatin frequently causes a dose-dependent sensory neuropathy that often but not always resolves following the cessation of therapy. FOLFIRI and FOLFOX are equal in efficacy. In metastatic disease, these regimens may produce median survivals of 2 years.

Monoclonal antibodies are also effective in patients with advanced colorectal cancer. Cetuximab (Erbitux) and panitumumab (Vectibix) are directed against the epidermal growth factor receptor (EGFR), a transmembrane glycoprotein involved in signaling pathways affecting growth and proliferation of tumor cells. Both cetuximab and panitumumab, when given alone, have been shown to benefit a small proportion of previously treated patients, and cetuximab appears to have therapeutic synergy with such chemotherapeutic agents as irinotecan, even in patients previously resistant to this drug; this suggests that cetuximab can reverse cellular resistance to cytotoxic chemotherapy. The antibodies are not effective in the ~65% subset of colon tumors that contain mutations in *ras* or *b-raf* genes and appear to be less likely to prove beneficial in the treatment of tumors arising from the right rather than left colon. The use of both cetuximab and panitumumab can lead to an acne-like rash, with the development and severity of the rash being correlated with the likelihood of antitumor efficacy. Inhibitors of the EGFR tyrosine kinase such as erlotinib (Tarceva) or sunitinib (Sutent) do not appear to be effective in colorectal cancer.

Bevacizumab (Avastin) is a monoclonal antibody directed against the vascular endothelial growth factor (VEGF) and is thought to act as an antiangiogenesis agent. The addition of bevacizumab to irinotecan-containing combinations and to FOLFOX appears to significantly improve the outcome observed with chemotherapy alone. The use of bevacizumab can lead to hypertension, proteinuria, and an increased likelihood of thromboembolic events.

Emerging data suggest that the use of checkpoint inhibitors (i.e., PD-1 and PD-2) as immunotherapy is more effective than chemotherapy in the subset (15%) of patients with metastatic colorectal cancer whose tumors are mismatch repair protein deficient (i.e., microsatellite unstable). Patients with solitary hepatic metastases without clinical or radiographic evidence of additional tumor involvement should be considered for partial liver resection, because such procedures are associated with 5-year survival rates of 25–30% when performed on selected individuals by experienced surgeons.

The administration of 5-FU and LV for 6 months after resection of tumor in patients with stage III disease leads to a 40% decrease in recurrence rates and 30% improvement in survival. The likelihood of recurrence has been further reduced when oxaliplatin has been combined with 5-FU and LV (e.g., FOLFOX). Reducing the duration of such oxaliplatin-containing therapy from 6 months to 3 months in patients with less invasive tumors (T_{1–2}, N₁) has been shown to result in a similar therapeutic benefit with reduced side effects (i.e., neurotoxicity) whereas 6 months of such therapy continues to be recommended for optimally treating patients with more advanced stage III tumors (T_{3–4} and/or N₂). Unexpectedly, the addition of irinotecan to 5-FU and LV as well as the addition of either bevacizumab or cetuximab to FOLFOX did not significantly enhance outcome.

Patients with stage II tumors do not appear to benefit appreciably from adjuvant therapy, with the use of such treatment generally restricted to those patients having biologic characteristics (e.g., perforated tumors, T4 lesions, lymphovascular invasion) that place them at higher likelihood for recurrence.

In rectal cancer, the delivery of preoperative or postoperative combined-modality therapy (5-FU or capecitabine plus radiation therapy) reduces the risk of recurrence and increases the chance of cure for patients with stage II and III tumors, with the preoperative approach being better tolerated.

CANCERS OF THE ANUS

Cancers of the anus account for 1–2% of the malignant tumors of the large bowel. Most such lesions arise in the anal canal, the anatomic area extending from the anorectal ring to a zone approximately halfway between the pectinate (or dentate) line and the anal verge. Carcinomas arising proximal to the pectinate line (i.e., in the transitional zone between the glandular mucosa of the rectum and the squamous epithelium of the distal anus) are known as *basaloid*, *cuboidal*, or *cloacogenic* tumors; about one-third of anal cancers have this histologic pattern. Malignancies arising distal to the pectinate line have squamous histology, ulcerate more frequently, and constitute ~55% of anal cancers. The prognosis for patients with basaloid and squamous cell cancers of the anus is identical when corrected for tumor size and the presence or absence of nodal spread.

The development of anal cancer is associated with infection by human papillomavirus, the same organism etiologically linked to cervical and oro-pharyngeal cancers. The virus is sexually transmitted. The infection may lead to anal warts (*condyloma acuminata*), which may progress to anal intraepithelial neoplasia and on to squamous cell carcinoma. The risk for anal cancer is increased among homosexual males, presumably related to anal intercourse. Anal cancer risk is increased in both men and women with AIDS, possibly because their immunosuppressed state permits more severe papillomavirus infection. Vaccination against human papilloma viruses appears to reduce the eventual risk for anal cancer. Anal cancers occur most commonly in middle-aged persons and are more frequent in women than men. At diagnosis, patients may experience bleeding, pain, sensation of a perianal mass, and pruritus.

Radical surgery (abdominal-perineal resection with lymph node sampling and a permanent colostomy) was once the treatment of choice for this tumor type. The 5-year survival rate after such a procedure was 55–70% in the absence of spread to regional lymph nodes and <20% if nodal involvement was present. An alternative therapeutic approach combining external beam radiation therapy with concomitant chemotherapy (5-FU and mitomycin C) has resulted in biopsy-proven disappearance of all tumor in >80% of patients whose initial lesion was <3 cm in size. Tumor recurrences develop in <10% of these patients, meaning that ~70% of patients with anal cancers can be cured with nonoperative treatment and without the need for a colostomy. Surgery should be reserved for the minority of individuals who are found to have residual tumor after being managed initially with radiation therapy combined with chemotherapy. The use of checkpoint immunotherapy (i.e., PD-1 inhibition) has been beneficial in some patients with recurrent disease.

FURTHER READING

- ANDRÉ T et al: Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. *N Engl J Med* 383:2207, 2020.
- COLÓN-LÓPEZ V et al: Anal cancer risk among people with HIV infection in the United States. *J Clin Oncol* 36:68, 2018.
- DEKKER E et al: Colorectal cancer. *Lancet* 394:1467, 2019.
- GROTHEY A et al: Duration of adjuvant chemotherapy for stage III colon cancer. *N Engl J Med* 378:1177, 2019.
- INADOMI JM: Screening for colorectal neoplasia. *N Engl J Med* 376:149, 2017.
- MARTIN D et al: Anal squamous cell carcinoma—state of the art management and future perspectives. *Cancer Treat Rev* 65:11, 2018.

PETRELLI F et al: Prognostic survival associated with left-sided vs right-sided colon cancer. A systemic review and meta-analysis. *JAMA Oncol* 3:211, 2017.

SALEM ME et al: Evaluation of the change of outcomes over a 10-year period in patients with stage III colon cancer: Pooled analysis of 6501 patients treated with fluorouracil, leucovorin, and oxaliplatin in the ACCENT database. *Ann of Oncol* 31:480, 2020.

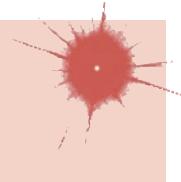
SCLAFANI F et al: Systemic therapies for advanced squamous cell anal cancer. *Curr Oncol Rep* 20:53, 2018.

SIEGEL RL et al: Colorectal cancer statistics 2020. *CA Cancer J Clin* 70:145, 2020.

82

Tumors of the Liver and Biliary Tree

Josep M. Llovet



Liver cancer is the sixth most common cancer worldwide, the fourth leading cause of cancer-related deaths, and the leading cause of death among cirrhotic patients. Liver cancer comprises a heterogeneous group of malignant tumors with different histologic features and unfavorable prognosis that range from hepatocellular carcinoma (HCC; 85–90% cases), intrahepatic cholangiocarcinoma (iCCA; 10%), and other malignancies accounting for <1% of tumors, such as fibrolamellar HCC, mixed HCC-iCCA, epithelioid hemangioendothelioma, and the pediatric cancer hepatoblastoma. The burden of liver cancer is increasing globally in almost all countries, and it is estimated to reach 1 million cases by 2025.

HEPATOCELLULAR CARCINOMA

■ EPIDEMIOLOGY AND RISK FACTORS

Overall, liver cancer accounts for 7% of all cancers (~850,000 new cases each year), and HCC represents 90% of primary liver cancers. The highest incidence rates of HCC occur in Asia and sub-Saharan Africa due to the high prevalence of hepatitis B virus (HBV) infection, with 20–35 cases per 100,000 inhabitants. Southern Europe and North America have intermediate incidence rates (10 cases per 100,000), whereas Northern and Western Europe have low incidence rates of <5 cases per 100,000 inhabitants. In the United States, liver cancer is ranked number one in terms of increased mortality during the past two decades, with an incidence of 35,000 cases per year (Fig. 82-1). HCC has a strong male preponderance with a male-to-female ratio estimated to be 2.5. The incidence increases with age, reaching a peak at 65–70 years old. In Chinese and in black African populations (where vertical transmission of HBV occurs), the mean age is 40–50 years. By contrast, in Japan, mean age in men is now around 75 years.

The risk factors for HCC are well established (Fig. 82-1). The main risk factor is cirrhosis—and associated chronic liver damage caused by inflammation and fibrosis—of any etiology, which underlies 80% of HCC cases worldwide and results from chronic infection by HBV or hepatitis C virus (HCV) infection, alcohol abuse, metabolic syndrome, and hemochromatosis (associated with *HFE1* gene germline mutations). Cirrhotic patients represent 1% of the human population, and one-third of them will develop HCC during their lifetime. Long-term follow-up studies have established an annual risk of HCC development of 3–8% in HBV- or HCV-infected cirrhotic patients. HCC is less common (1–3% per year) in cirrhosis associated with alcohol, nonalcoholic steatohepatitis (NASH), α_1 -antitrypsin deficiency, autoimmune hepatitis, Wilson's disease, and cholestatic liver disorders. Predictors of liver cancer development among cirrhotic patients have been associated with liver disease severity (platelet count of <100,000/ μ L, presence

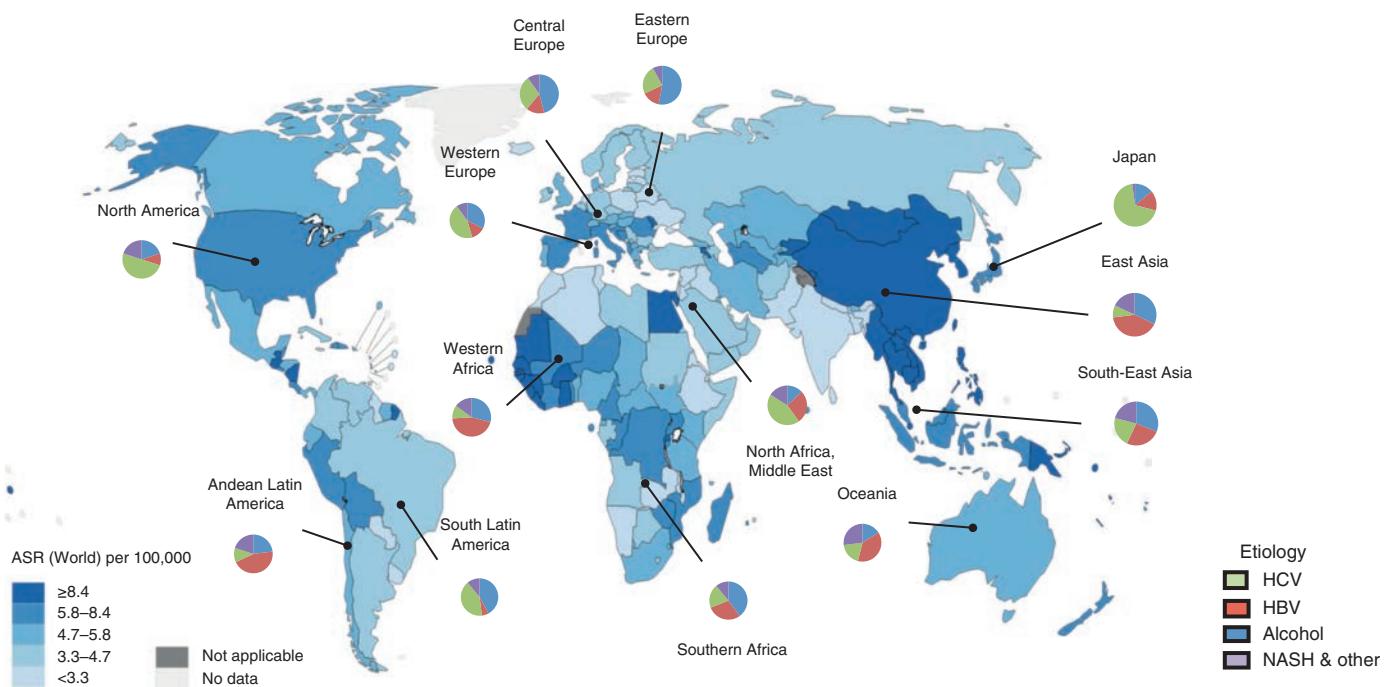


FIGURE 82-1 Distribution of hepatocellular carcinoma incidence according to geographical area and etiology. HBV, hepatitis B virus; HCV, hepatitis C virus; NASH, nonalcoholic steatohepatitis. (Reproduced with permission from JM Llovet et al: Hepatocellular carcinoma. Nat Rev Disease Primers 6:7, 2021.)

of portal hypertension), the degree of liver stiffness as measured by transient elastography, and liver gene signatures capturing the *cancer field effect*.

In terms of attributable risk fraction, HBV infection—a DNA virus that can cause insertional mutagenesis and affects 400 million people globally—accounts for ~60% of HCC cases in Asia and Africa and 20% in the Western world. Among patients with HBV infection, a family history of HCC, HBeAg seropositivity, high viral load, and genotype C are independent predictors of HCC development. Chronic treatments with effective antiviral HBV therapies are able to significantly decrease the risk of cancer. HCV infection—an RNA virus that affects 170 million people—is responsible for ~30% of cases and is the main cause of HCC in Europe and North America. Among patients with HCV infection, HCC occurs almost exclusively when relevant liver damage is present (either advanced fibrosis—Metavir F3 [Metavir is a scoring system for hepatic histology that grades fibrosis from 0 to 4 with higher numbers indicating more fibrosis]—or cirrhosis), particularly if associated with HCV genotype 1b. In addition, a polymorphism that activates EGFR, the epidermal growth factor receptor, is associated with HCV-HCC in several studies. Antiviral therapies with interferon regimens are able to prevent cirrhosis development and HCC occurrence. Direct-acting antiviral agents (DAA) induce sustained virologic response, i.e., clearance of HCV infection, in most of cases, thus resulting in 50–80% reduction in HCC risk.

Alcohol consumption and metabolic syndrome due to diabetes and obesity are responsible for ~30% of cases. NASH is becoming the leading cause of cirrhosis in developed countries and currently represents ~15–20% of HCC cases in the West. The annual incidence of HCC in NASH-related cirrhosis (1–2%/year) justifies including patients at risk in surveillance programs. Nonetheless, it has to be taken into account that 25–30% of NASH-associated HCC occurs in the absence of cirrhosis. A PNPLA3 polymorphism is strongly associated with fatty and alcoholic chronic liver diseases and HCC occurrence. Other cofactors contributing to HCC development are tobacco and aflatoxin B1, a fungal carcinogen present in food supplies that induces TP53 mutations. Finally, infection with adeno-associated virus 2 is associated with HCC in individuals without cirrhosis. Aside from the associations described above, genome-wide association studies have not yet confirmed polymorphisms predisposing to HCC development.

■ MOLECULAR PATHOGENESIS

HCC development is a complex multistep process that starts with precancerous cirrhotic nodules, so-called low-grade dysplastic nodules (LGNs) that evolve to high-grade dysplastic nodules (HGNs) that can transform into early-stage HCC. Molecular studies support the pivotal role of adult hepatocytes as the cell of origin, either by directly transforming to HCC or by de-differentiating into hepatocyte precursor cells. Alternatively, progenitor cells also give rise to HCC with progenitor markers.

Genomic analysis has provided a clear picture of the main drivers responsible for HCC initiation and progression. This tumor results from the accumulation of around 40–60 somatic genomic alterations per tumor, of which 4–8 are considered driver cancer genes. HCC is a prototypical inflammation-associated cancer, where immune microenvironment and oxidative stress present in chronically damaged livers play pivotal roles in inducing mutations. In preneoplastic HGN, mutations in telomere reverse transcriptase (*TERT*) gene (20% of cases) and gains in 8q have been described. Oncogenic transformation occurs upon additional genomic hits including Wnt/β-catenin pathway activation, reexpression of fetal genes, deregulation of protein folding machinery, and the response to oxidative stress. Genomic studies and next-generation sequencing conducted during the past decade have enabled a description of the landscape of mutations, signaling pathways, and molecular classification of the disease. Nonetheless, none of these data have yet translated into actual clinical benefits for any specific molecularly based subgroups.

Molecular Drivers The landscape of mutational drivers in HCC identified by deep-genome sequencing is detailed in **Table 82-1**. The most common mutations are in the *TERT* promoter (56%), *TP53* (27%), *CTNNB1* (26%), *ARID2* (7%), *ARID1A* (6%), and *AXIN1* (5%) genes. These mutated genes participate in cell-cycle control and senescence (*TERT* and *TP53*), cell differentiation (*CTNNB1* and *AXIN1*), and chromatin remodeling (*ARID2* and *ARID1A*). Genes commonly mutated in other solid tumors such as *EGFR*, *HER2*, *PIK3CA*, *BRAF*, or *KRAS* are rarely mutated in HCC (<5%). Overall, only ~20–25% of HCCs have at least one actionable mutation. Some risk factors have been associated with specific molecular aberrations. HBV integrates into the genome of driver genes, such as the *TERT* promoter, *MLL4*, and cyclin E1 (*CCNE1*). Alcohol abuse and HCV infection have been

TABLE 82-1 Molecular Aberrations Common in Hepatocellular Carcinoma (HCC)^a

PATHWAY	TARGET	PREVALENCE (%)
Mutations		
Telomere stability	<i>TERT</i> promoter	56
p53/cell-cycle control	<i>TP53</i>	27
	<i>ATM</i>	3
	<i>RB1</i>	3
Wnt/β-catenin signaling	<i>CTNNB1</i>	26
	<i>AXIN1</i>	5
Chromatin remodeling	<i>ARID1A</i>	6
	<i>ARID2</i>	7
	<i>KMT2A</i>	3
	<i>KMT2C</i>	3
Ras/PI3K/mTOR pathway	<i>RPS6KA3</i>	3
	<i>TSC1/TSC2</i>	3
Oxidative stress	<i>NFE2L2</i>	3
	<i>KEAP1</i>	3
High-level focal amplifications		
VEGF signaling	<i>VEGFA</i>	3
FGF signaling	<i>FGF19</i>	6
Cell-cycle control	<i>CCND1</i> protein	7
Target with homozygous deletion		
TP53/cell-cycle control	<i>CDKN2A</i>	5
	<i>TP53</i>	4
	Retinoblastoma 1	4
Wnt/β-catenin signaling	<i>AXIN1</i>	3

^aRecurrent mutations, focal amplifications, or homozygous deletions in HCC based on next-generation sequencing analyses.

associated with *CTNNB1* mutations. *TP53* mutations are the most frequent alterations with a specific hotspot of mutation (R249S) in patients with aflatoxin B1 exposure.

Studies assessing copy number alterations in HCCs have consistently identified: (1) high-level amplifications at 5–10% prevalence containing oncogenes in 11q13 (*CCND1* and *FGF19*) and 6p21 (*VEGFA*), *TERT* focal amplification, and homozygous deletion of *CDKN2A*; and (2) common amplifications containing *MYC* (8q gain) and *MET* genes (focal gains of 7q31). Activation of the FGF19-FGFR4 pathway mediated by epigenetic mechanisms (~20%) or high-level amplifications of 11q13 (6%) or *VEGFA* gains (high-level gains of 6p21) are also potential therapeutic targets.

Signaling Pathways Several signaling pathways have been implicated in HCC progression and dissemination. Activation of these pathways can result from structural alterations (mutations and amplifications/losses) or epigenetic modifications. In brief, (1) *TERT* overexpression occurs in 90% of cases, particularly related to promoter *TERT* mutations or amplifications; (2) inactivation of p53 and alterations of cell cycle are major defects in HCC, particularly in cases related to HBV infection; (3) Wnt/β-catenin pathway activation occurs in 50% of cases, either as a result of β-catenin or *AXIN1* mutation or overexpression of Frizzled receptors or inactivation of E-cadherin; (4) PI3K/PTEN/Akt/mTOR pathway is activated in 40–50% of HCCs due to mutation and focal deletion of the tuberous sclerosis complex (*TSC1*/*TSC2*) genes, *PTEN*, or ligand overexpression of EGF or insulin-like growth factor (IGF) upstream signals; (5) Ras MAPK signaling is activated in half of early and almost all advanced HCCs, and activation results from upstream signaling by EGF, IGF, and MET activation; (6) insulin-like growth factor receptor (IGFR) signaling is activated in 20% of cases through overexpression of the oncogenic ligand IGF2; (7) dysregulation of the c-MET receptor and its ligand HGF, critical for hepatocyte regeneration after liver injury, is a common event in advanced HCC (50%); (8) vascular endothelial growth factor (VEGF) signaling is the cornerstone of angiogenesis in HCC, along with activated angiogenic

pathways such as Ang2 and FGF signaling; and (9) chromatin remodeling complexes and epigenetic regulators are frequently altered in HCC due to *ARID1A* and *ARID2* mutations.

Molecular and Immune Classes Genomic studies have revealed two molecular subclasses of HCC, each representing ~50% of patients. The proliferative subclass is enriched by activation of Ras, mTOR, and IGF signaling and *FGF19* amplification and is associated with HBV-related etiologies, overexpression of α-fetoprotein, and poor outcomes. By contrast, the so-called nonproliferative subclass contains a subtype characterized by *CTNNB1* mutations and better outcome. Another classification based on immune status has been proposed. It defines an immune HCC class in ~25% of cases characterized by immune infiltrate with expression of PD-1/PD-L1, enrichment of T cell activation, and better outcome and an immune excluded class with activation of pathways related with immune escape (i.e., Wnt signaling) or absence of T cell infiltrate. This excluded class has been proposed to be associated with resistance to immune checkpoint inhibitors, although direct translation of molecular subclasses into clinical decision making has yet to be achieved.

■ PREVENTION AND EARLY DETECTION

Prevention Primary prevention of HCC can be achieved by vaccination against HBV and effective treatment of HBV and HCV infection. Studies assessing the impact of universal vaccination against HBV infection have reported a significant decrease of the incidence of HCC. HBV vaccination is recommended to all newborns and high-risk groups, following World Health Organization guidelines. Vaccination is also recommended in people with risk factors for acquiring HBV infection, such as health workers, travelers to areas where HBV infection is prevalent, injection drug users, and people with multiple sex partners.

Effective antiviral treatments for patients with chronic HBV infection—achieving undetectable viral titers (circulating HBV-DNA)—reduce the risk of HCC development. Evidence of this effect is supported by one randomized trial and several cohort studies. Treatment of HCV has dramatically advanced with the new DAAs, which yield >90% sustained virologic response (SVR) rates after 12 weeks of treatment. This effect has a direct implication in reducing HCC incidence in patients with cured chronic HCV infection. Once cirrhosis is established, the incidence of HCC is lower for patients with SVR than for those with active viral disease, although they continue to have persistent HCC risk (>1% per year). Additional putative chemopreventive agents have been proposed to reduce HCC incidence in at-risk populations. Aspirin is associated with HCC cumulative incidence reduction in large studies from 8% to 4%. Similarly, compelling cohort and case-control studies demonstrated a dose-dependent relationship between coffee consumption and reduced HCC incidence. As a result, coffee consumption is recommended as a chemoprevention strategy in patients with chronic liver disease.

Surveillance The aim of surveillance is to obtain a reduction in disease-related mortality. This is usually achieved through early detection that enhances the applicability and cost-effectiveness of curative therapies. U.S. and European guidelines recommend surveillance for patients at high risk for HCC on the basis of cost-effectiveness analyses. As a general rule, high-risk populations are considered those presenting an incidence cutoff >1.5% for patients with cirrhosis and 0.2% for patients with chronic hepatitis B. However, the strength of evidence supporting surveillance is modest and is based on two randomized studies conducted in China and a meta-analysis of observational studies. Overall, these studies conclude that surveillance identifies patients with smaller tumors who are more likely to undergo curative procedures. Because of lead time bias and length time bias, it cannot be concluded that surveillance ultimately reduces HCC-related mortality.

Surveillance is recommended for cirrhotic patients due to any cause, those with HCV-related advanced fibrosis (Metavir score of F3), and patients with chronic HBV infection if Asian and aged >40 years, if African and aged >20 years, if there is a family history of HCC, or if the

patient has sufficient risk by risk scores such as PAGE-B. In terms of liver dysfunction, the presence of advanced cirrhosis (Child-Pugh class C) prevents potentially curative therapies from being employed, and thus surveillance is not recommended. As an exception, patients on the waiting list for liver transplantation, regardless of liver functional status, should be screened for HCC in order to detect tumors exceeding conventional criteria and to define priority policies for transplantation. Complex scoring systems to identify at-risk populations are not yet recommended by guidelines.

Ultrasonography every 6 months, with or without serum α fetoprotein (AFP) levels, is the recommended method of surveillance. It has a sensitivity of 65–80% and a specificity of >90% for early detection. A 3-month interval does not enhance outcomes, and survival is lower with 12-month compared with 6-month intervals. A shorter follow-up interval (every 3–4 months) is recommended when a nodule of <1 cm has been detected. Computed tomography (CT) and magnetic resonance imaging (MRI) are not recommended as screening tools due to lack of data on accuracy, high cost, and possible harm (i.e., radiation with CT). Exceptionally, these techniques can be considered in patients with obesity and fatty liver, where visualization with ultrasound is difficult. Accurate tumor biomarkers for early detection need to be developed. Use of AFP levels as a stand-alone method identifies patients with HCC with 60% sensitivity but has high false-positive results. One main limitation of AFP is that only a small proportion of early tumors (~20%) present with abnormal AFP serum levels. Combining AFP with ultrasound might increase the HCC detection rate from 8–30% compared to ultrasound and depending on the performance by experienced personnel as a stand-alone method. The accuracy of other serum biomarkers proposed, such as des- γ -carboxyprothrombin (DCP) and the L3 fraction of AFP (AFP-L3), in early detection is not known.

Despite the fact that surveillance is cost-effective in HCC, the global implementation of such programs is estimated to engage ~50% of the target population in Europe and ~30% in the United States. Public health policies encouraging the implementation of such programs should lead to an increase in early tumor detection.

Diagnosis HCC is generally diagnosed at early or intermediate stages in Western countries but at advanced stages in most Asian (except Japan) and African countries. A surveillance program yields early diagnosis in 70–80% of cases. At these stages, the tumor is asymptomatic, and diagnosis can be made by noninvasive (radiologic) or invasive (biopsy) approaches. Without surveillance, HCC is discovered either as a radiologic finding or due to cancer-related symptoms. If symptoms are present, the disease is already at an advanced stage with a median life expectancy <1 year. Symptoms include malaise, weight loss, anorexia, abdominal discomfort, or signs related to advanced liver dysfunction.

NONINVASIVE (RADIOLOGIC) DIAGNOSIS Patients enrolled in a surveillance program are diagnosed by identification of a new liver nodule on abdominal ultrasound. Noninvasive diagnostic criteria can only be applied to cirrhotic patients and are based on imaging techniques obtained by four-phase multidetector CT scan (four phases are unenhanced, arterial, venous, and delayed) or dynamic contrast-enhanced MRI. A flowchart of diagnosis and recall policy recommended by U.S. and European guidelines is summarized in Fig. 82-2. Radiologic diagnosis is achieved with a high degree of confidence if the lesion is ≥ 1 cm in diameter and shows the *radiologic hallmarks of HCC* by one imaging technique. Using contrast-enhanced imaging techniques, the typical hallmark of HCC consists of vascular uptake of the nodule in the arterial phase with washout in the portal venous or delayed phases. This radiologic pattern captures the hypervascular nature characteristic of

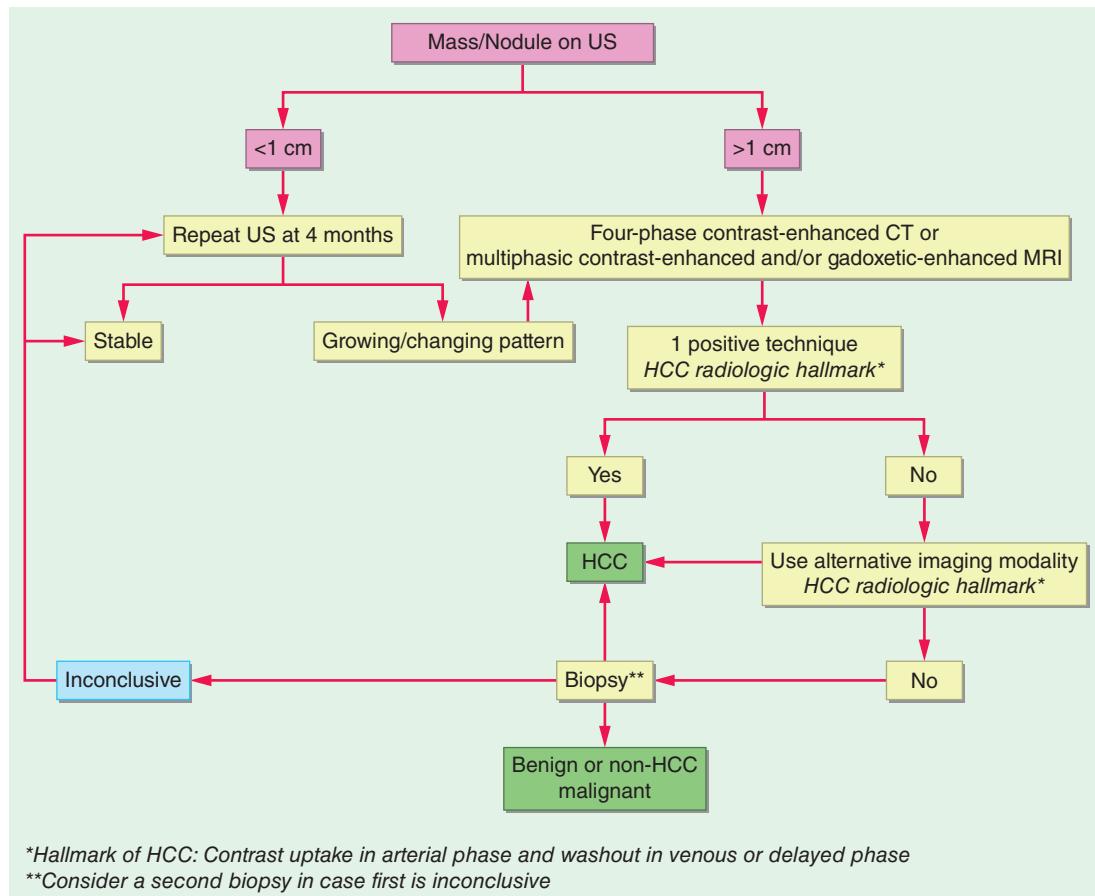


FIGURE 82-2 Recall diagnosis schedule for hepatocellular carcinoma (HCC) from the European Association for the Study of Liver Disease (EASL). **Pink color:** Size of the tumor at the time of detection by ultrasound (US). **Yellow color:** If a nodule of <1 cm is detected, repeated US at 4 months is recommended. If a nodule of >1 cm is detected, CT or MRI will be performed. Presence of *radiological hallmarks of HCC* by one imaging technique will suffice for diagnosis. This might require using one or two imaging techniques. If no diagnosis is established, then tissue biopsy would be recommended. **Green color:** Final diagnosis could be either HCC, benign tumor or non-HCC malignant. **Blue color:** If after 2 biopsies the is no conclusive diagnosis then consider follow-up with US at 4 months. (Reproduced with permission from European Association for the Study of the Liver: EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatology* 69:182-236, 2018.)

HCC. In these scenarios, the diagnostic specificity is ~95–100% and a biopsy is not necessary. Nodules <1 cm in size are unlikely to be HCC and would be very difficult to diagnose; thus, ultrasound follow-up at 3–4 months is recommended. MRI with liver-specific contrast agents is accepted as a diagnostic tool (Fig. 82-2). Contrast-enhanced ultrasound and angiography are less accurate for HCC diagnosis. Positron emission tomography (PET) scan performs poorly for early diagnosis. AFP levels ≥400 ng/dL are highly suspicious, but not diagnostic, of HCC according to guidelines.

The Liver Imaging Reporting and Data System (LI-RADS) has been proposed as a way of classifying radiologic findings. Essentially, nodules >10 mm visible on multiphase exams are assigned category codes reflecting their relative probability of being benign, HCC, or other hepatic malignant neoplasms. LI-RADS-1 lesions have a 0% probability of HCC, whereas lesions assigned to the LI-RADS-5 category have a 96% probability of HCC. LI-RADS-M category comprises lesions with malignant radiologic features but are not HCC malignancies in >50% of cases.

PATHOLOGIC DIAGNOSIS Pathologic diagnosis is required in two scenarios: (1) in patients without cirrhosis and (2) if imaging is not typical in at least one of two imaging techniques (CT and MRI). This occurs mainly with early-stage HCC lesions. Biopsy has not been used as the gold standard in clinical practice because of variation introduced by sampling and complications. Nonetheless, with the advent of molecular therapies and precision oncology, some guidelines advocate obtaining tissue samples in the setting of all research studies in HCC, even if radiologic criteria are met. Sensitivity of liver biopsies ranges between 70 and 90% for all tumor sizes but decreases to <50% in tumors 1–2 cm in size. The risk of complications such as tumor seeding and bleeding after liver biopsy is ~3%. Biopsies should be assessed by an expert

hepatopathologist. The use of special stains may help to resolve diagnostic uncertainties. Positive staining in two of four markers (glypican 3 [GPC3], glutamine synthetase, heat shock protein 70 [HSP70], and clathrin heavy chain) is highly specific for HCC. Gene expression blueprints (glypican 3, LYVE1, and survivin) are also able to differentiate HGDNs from early HCC. Additional staining can be considered to detect progenitor cell features (K19 and EpCAM) or assess neovascularization (CD34). A negative biopsy does not eliminate the diagnosis of HCC. A second biopsy is recommended in case of inconclusive findings or if growth or change in enhancement pattern is identified during follow-up (Fig. 82-2).

TREATMENT

Overview The landscape of management of HCC has substantially changed during the past decade. Several treatments have been adopted as standard of care according to clinical practice guidelines. For early stages, resection, liver transplantation, and local ablation have substantially improved life expectancy, with median overall survival (OS) times beyond 5 years (Fig. 82-3). For intermediate stages, transarterial chemoembolization (TACE) has improved OS from 16 months (natural history) to 20–30 months. Finally, systemic drugs for advanced tumors (atezolizumab plus bevacizumab, sorafenib, lenvatinib, regorafenib, cabozantinib, and ramucirumab) have improved median survival times from 8 months to 19 months in first-line settings and 10 months in second-line settings. Currently, several unmet needs, such as adjuvant therapies after resection or local ablation and improving outcomes at intermediate/advanced stages with combination therapies including immunotherapies, are being addressed in the setting of phase 3 investigations (Fig. 82-3).

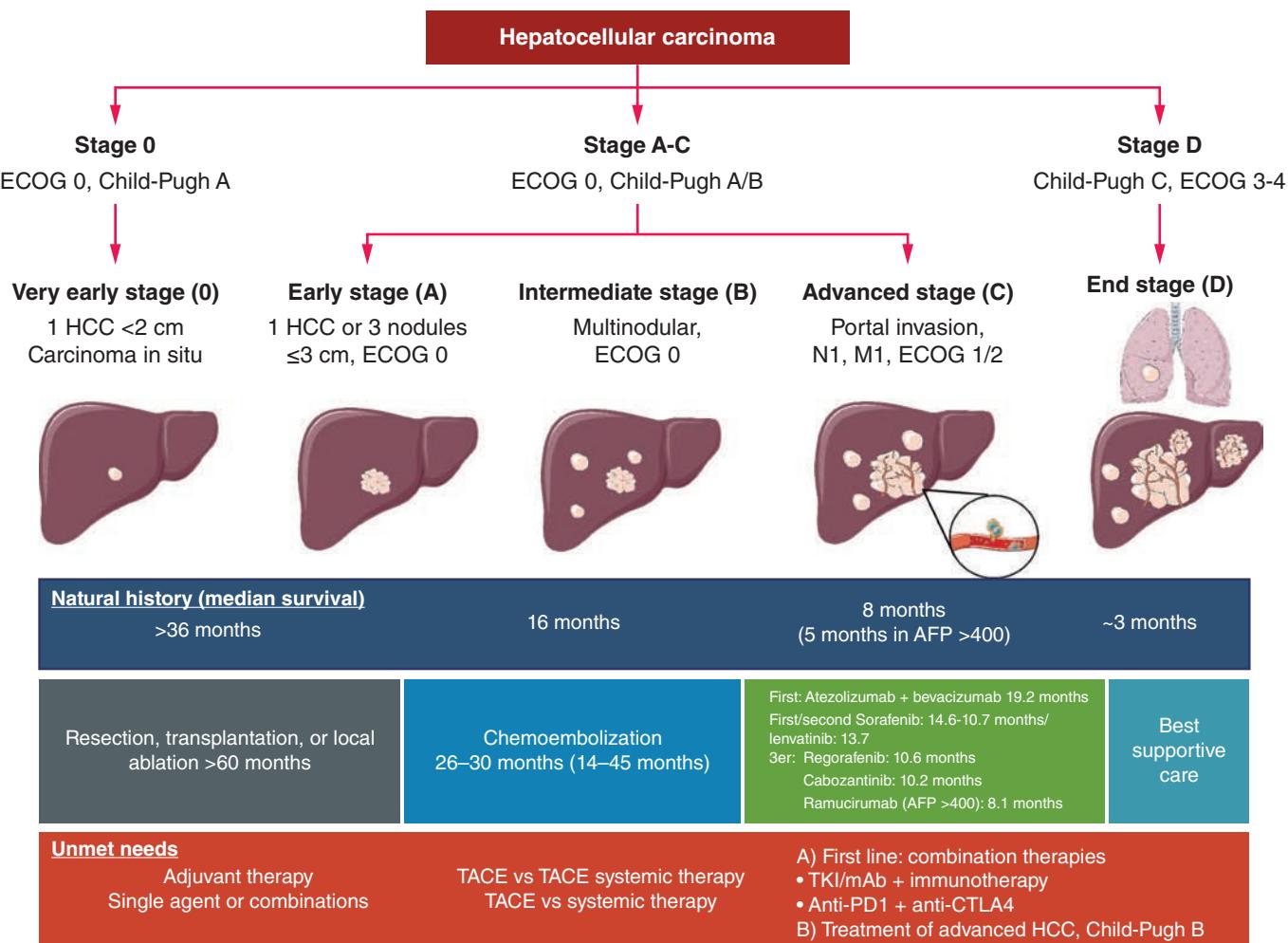


FIGURE 82-3 Natural history, impact of therapies, and unmet needs in hepatocellular carcinoma (HCC). AFP, α -fetoprotein; ECOG, Eastern Cooperative Oncology Group performance status; mAb, monoclonal antibody; TACE, transarterial chemoembolization; TKI, tyrosine kinase inhibitor.

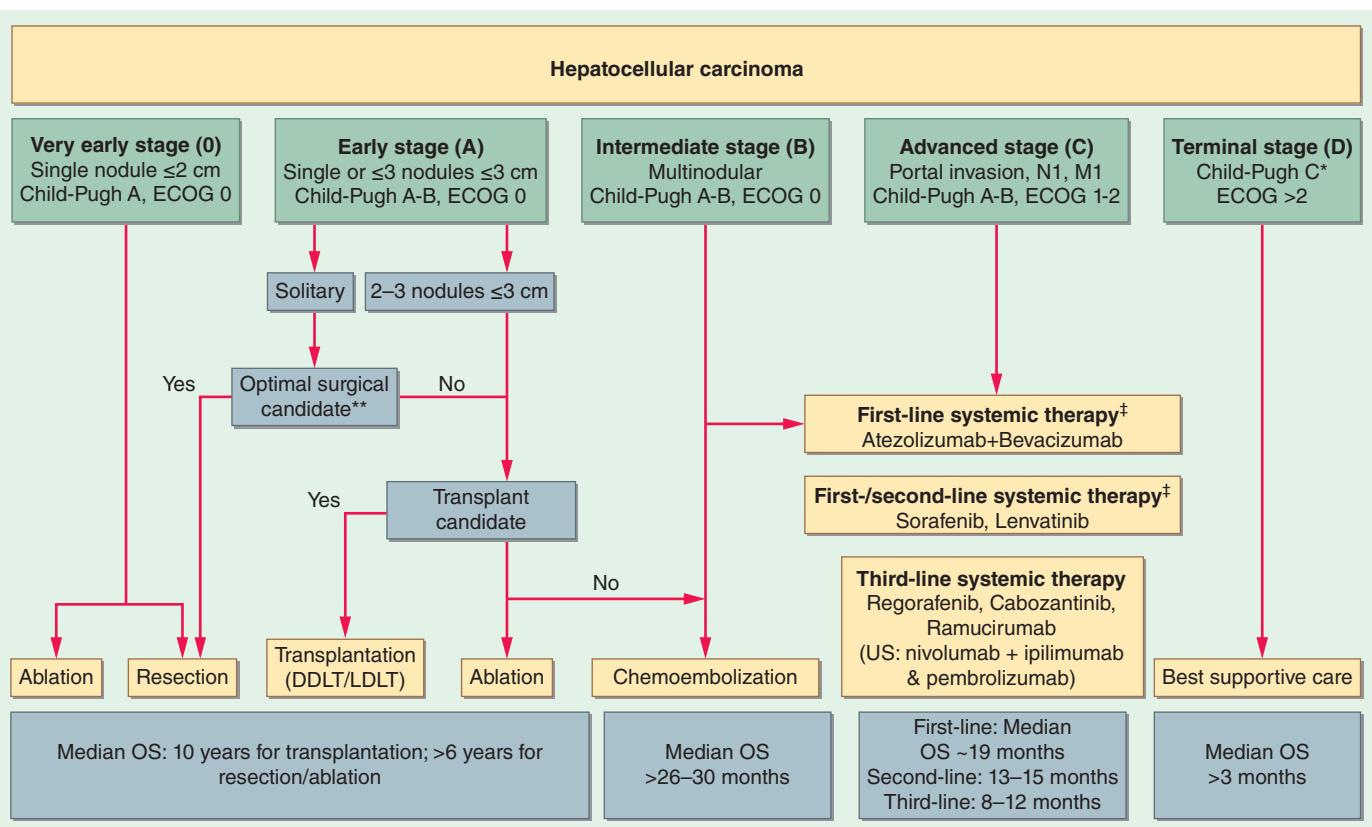


FIGURE 82-4 Staging system and therapeutic strategy. Barcelona Clinic Liver Cancer (BCLC) classification comprises five stages that select the best candidates for therapies according to evidence-based data. Patients with asymptomatic early tumors (stages 0–A) are candidates for radical therapies (resection, transplantation, or local ablation). Asymptomatic patients with multinodular hepatocellular carcinoma (HCC) (stage B) are suitable for transcatheter arterial chemoembolization (TACE), whereas patients with advanced symptomatic tumors and/or an invasive tumoral pattern (stage C) are candidates to receive systemic therapies. End-stage disease (stage D) includes patients with poor prognosis who should be treated by best supportive care. *Patients with end-stage liver disease if Child-Pugh class C should first be considered for liver transplantation. ‡Atezolizumab plus bevacizumab has been approved as new first-line treatment for advanced HCC. Nonetheless, sorafenib and lenvatinib are still considered first line options when there is a contraindication for the combination treatment. DDLT, deceased donor liver transplantation; ECOG, Eastern Cooperative Oncology Group performance status; LDLT, living donor liver transplantation; OS, overall survival. (Reproduced with permission from JM Llovet et al: Trial Design and Endpoints in Hepatocellular Carcinoma: AASLD Consensus Conference 73:158, 2021.)

Staging Systems and Treatment Allocation Staging systems are aimed at stratifying patients according to prognostic factors and outcome and allocating the best available therapies according to evidence. The most accepted staging system is the Barcelona Clinic Liver Cancer (BCLC) Classification, which is endorsed by U.S. and European clinical practice guidelines (Fig. 82-4). This staging system defines five prognostic subclasses and allocates specific treatments for each stage. The BCLC staging system has been externally validated by numerous studies. It is an evolving system that allows incorporation of new therapies and treatment-dependent variables as new evidence emerges. Ten treatments have been shown to improve survival in HCC and thus have been incorporated in the therapeutic algorithm: surgical resection, liver transplantation, radiofrequency (RF) ablation, chemoembolization, and systemic therapies (atezolizumab-bevacizumab, sorafenib, lenvatinib, regorafenib, cabozantinib, and ramucirumab). The BCLC assigns each patient to a given treatment allocation. Treatment stage migration is also applied by this scheme, meaning that if patients are not candidates for the selected therapy, the next effective therapy at more advanced stages can be given.

In HCC, three parameters are relevant for defining treatment strategy: tumor status, cancer-related symptoms, and liver dysfunction. The BCLC staging captures all three variables and allocates patients to treatments according to evidence. Since >80% of patients have two diseases, HCC and cirrhosis, a clear measurement of liver dysfunction should be in place. The prognosis of chronic liver disease is commonly assessed using the Child-Pugh score, which uses five clinical measures—total bilirubin, serum albumin, prothrombin time, ascites severity, and hepatic encephalopathy grade—to classify patients into one of three groups (A–C) of predicted survival rates. In brief, Child-Pugh class A reflects well-preserved liver function, Child-Pugh class B indicates moderate liver dysfunction with a median life expectancy of ~3 years, and Child-Pugh class C indicates severe liver dysfunction with life

expectancy of ~1 year. At early BCLC stages, more granular criteria to define patients with very-well-preserved liver function (Child-Pugh hyper-A class; those patients with normal bilirubin and without portal hypertension) need to be in place to select candidates for resection. Modifications of Child-Pugh scoring or the Model for End-Stage Liver Disease (MELD) score have not been adopted for treatment allocation, except for prioritization on the waiting list for liver transplantation (MELD score). The ALBI score, which is based only on serum albumin and bilirubin levels, has been shown to accurately stratify patients with HCC, particularly those with less severe liver dysfunction. Performance status is assessed using the Eastern Cooperative Oncology Group (ECOG) performance scale (a 5-point system where higher numbers indicate greater disability), and the presence of cancer-related symptoms (ECOG 1–2) is considered a sign of advanced stage. Patients with severe liver dysfunction (Child-Pugh class C) or performance status impairment (ECOG 3–4) are offered supportive care management.

Considering all of these prognostic and predictive variables and evidence-based treatment efficacy, five BCLC stages have been defined (Fig. 82-4). Patients with liver-only neoplastic disease, no symptoms (ECOG 0), and mild to moderate liver dysfunction (Child-Pugh A–B) can be classified as very early (stage 0) or early (stage A) or intermediate (stage B) stages depending on tumor size and number. Very early HCC (BCLC 0) is defined by single tumors ≤2 cm (if pathology is available, they should be well differentiated with absence of microvascular invasion or satellites). Early HCC (BCLC A) includes either single tumors or a maximum of three nodules of ≤3 cm in diameter. Intermediate stage (BCLC B) is defined by all other liver-only tumors. Conversely, HCC is considered at advanced stages (BCLC C) when patients present with cancer-related symptoms (ECOG 1–2) or tumors with macrovascular invasion (of any type, including branch, hepatic, or portal vein), lymph node involvement, or extrahepatic spread. Finally,

end-stage disease (BCLC D) is considered in cases of several impairment of quality of life/cancer-related symptoms (ECOG 3–4) or severe liver dysfunction (Child-Pugh C).

Around 40% of patients are diagnosed at stages 0 and A and, hence, are eligible for potentially curative therapies, resection, transplantation, or local ablation. These treatments provide median survival rates of 60 months and beyond, which are in sharp contrast with outcomes of 36 months reported in historical controls (Fig. 82-3, **Table 82-2**). No adjuvant therapy is recommended. Patients at intermediate stage (stage B) with preserved liver function have a documented natural history of around 16 months. These patients benefit from TACE as reported in two randomized studies and one meta-analysis and achieve an estimated survival of 25–30 months. None of the combination therapies with TACE have shown outcome advantages. Patients progressing on TACE or at advanced stage (stage C) benefit from systemic treatments. Sorafenib extends survival by ~3 months compared to placebo (from 7.9 to 10.7 months), whereas lenvatinib showed noninferiority compared to sorafenib (13.6 months vs 12.3 months, respectively). Atezolizumab (an anti-PD-L1 antibody) plus bevacizumab showed superiority compared to sorafenib (median survival 19.2 months vs 13.4 months). Three additional targeted therapies have shown improved survival compared to placebo in patients with HCC progressing on sorafenib: regorafenib, cabozantinib, and ramucirumab (only in patients with AFP >400 ng/mL). Therefore, these treatments have been adopted by guidelines and incorporated into the treatment algorithm. Patients with end-stage disease (BCLC D) should be considered for nutritional and psychological support and proper management of pain.

Although the BCLC establishes validated stages and treatment assignment according to evidence, clinical practice is not always aligned with this classification. In large cohort studies and surveys, only half of patients, or even less in Asia, are treated accordingly. Alternative

staging or scoring systems have been proposed, but none of them has acquired global consensus. In contrast to BCLC, some proposed systems capture the standard of practice in Asia, such as the Hong Kong classification or the Japan Integrated Staging score. These systems capture extended indications for resection and TACE applied in clinical practice in Asia. Finally, the tumor-node-metastasis (TNM) staging system is not used in HCC since it does not incorporate the main prognostic variables related to liver function and performance status.

Due to the complexities of HCC diagnosis and management, it is recommended that patients be sent to a referral center where all the armamentarium of therapies can be offered. In principle, patient management and outcome benefit from liver cancer multidisciplinary programs that include a hepatologist, oncologist, hepatobiliary and transplant surgeons, interventional and body imaging radiologist, hepatopathologist, and specialized nurses.

SURGICAL THERAPIES

Resection Surgical resection is the first-line option for noncirrhotic patients with early-stage HCC (BCLC 0 or A) with solitary tumors (Fig. 82-4). In cirrhotic patients, ablation competes with resection for BCLC 0 tumors (<2 cm in diameter). Which treatment is better is not defined. Cost-effectiveness approaches report a benefit for local ablation with RF. For single tumors >2 cm (BCLC A), resection remains the mainstay of treatment in patients with Child-Pugh hyper-A class, i.e., those patients with normal bilirubin and absence of portal hypertension (portal hypertension is defined by hepatic venous pressure gradient ≥10 mmHg). Surrogate measures of portal hypertension are presence of esophageal varices or platelet count <100,000/ μ L associated with splenomegaly. Anatomic resections following the functional segments of the liver are recommended to spare uninvolved

TABLE 82-2 Summary of Key Results of Randomized and Cohort Studies in the Management of Hepatocellular Carcinoma

TREATMENT OF EARLY- AND INTERMEDIATE-STAGE HCC				
TREATMENTS	HCC STAGE	TREATMENT ARMS	OUTCOMES (OS)	
Treatment for early HCC				
Resection	Early	Optimal (single nodule; no portal hypertension)	5 years: 50–70%	
		Suboptimal (multinodular or portal hypertension)	5 years: 35–55%	
Liver transplantation	Early	Milan (1 nodule ≤5 cm, 2–3 nodules ≤3 cm, no MVI, no EHS)	5 years: 70–80%	
	Early/intermediate	Downstaged (1 nodule ≤6.5 cm, ≤3 nodules ≤4.5 cm and total diameter ≤8 cm, no MVI, no EHS)	5 years: 60–70%	
Ablation	Early	RFA	Median: 50–60 months	
Treatments for intermediate HCC				
Transarterial therapies	Intermediate	TACE	Median: 20–32 months	
TREATMENT OF ADVANCED-STAGE HCC				
STUDY	TREATMENT	MEDIAN OS, MONTHS (HR, 95% CI)	MEDIAN PFS, MONTHS (HR, 95% CI)	ORR: MRECIST/RECIST
First-line therapies				
IMbrave150	Atezolizumab-bevacizumab	19.2 (HR 0.66, 0.52–0.85)	6.9 (HR 0.65, 0.53–0.81)	35.4%/29.8%
SHARP/REFLECT/ CheckMate-459	Sorafenib	10.7–14.6 (HR 0.69, 0.55–0.87)	3.7–3.8	2–9.2%/12.4%
REFLECT	Lenvatinib	13.6 (HR 0.92, 0.79–1.06)	7.4 (HR 0.66, 0.57–0.77)	24.1%/18.8%
Second-line therapies				
RESORCE	Regorafenib	10.6 (HR 0.63, 0.5–0.79)	3.1 (HR 0.46, 0.37–0.56)	11%/7%
CELESTIAL	Cabozantinib	10.2 (HR 0.76, 0.63–0.92)	5.2 (HR 0.44, 0.36–0.52)	NA/4%
REACH-2	Ramucirumab	8.5 (HR 0.71, 0.53–0.95)	2.8 (HR 0.45, 0.34–0.6)	NA/5%

Abbreviations: CI, confidence interval; EHS, extrahepatic spread; HCC, hepatocellular carcinoma; HR, hazard ratio; MRECIST, Modified Response Evaluation Criteria in Solid Tumors; MVI, microvascular invasion; NA, not available; NE, not evaluable; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; RECIST, Response Evaluation Criteria in Solid Tumors; RFA, radiofrequency ablation; TACE, transarterial chemoembolization.

liver parenchyma and to remove satellite tumors. Predictors of recurrence are tumor size and number and presence of microsatellites or microvascular invasion at the specimen analysis. Outcomes in suboptimal candidates lead to 5-year survival rates of ~35–55%, as opposed to 60–70% for ideal candidates (Table 82-2). Macrovascular invasion, extrahepatic involvement, and liver dysfunction (Child-Pugh B-C) are major contraindications for resection.

ADJUVANT TREATMENTS Tumor recurrence represents the major complication of resection (and local ablation) and occurs in 70% of cases at 5 years. Most recurrences are intrahepatic metastases, but at least one-third are considered de novo tumors, new clones developing in the cirrhotic carcinogenic field. The type of recurrence can only be defined by molecular studies. So far, no adjuvant therapies have been proven to improve outcome or prevent recurrence after resection/ablation. Randomized trials testing adjuvant sorafenib, retinoids, chemotherapies, or chemoembolization have been negative, and thus, no adjuvant therapy recommendation has been established for patients after resection or local ablation.

Liver Transplantation Liver transplantation is the first treatment choice for cirrhotic patients with single tumors ≤5 cm and portal hypertension (including Child-Pugh B and C) or with small multinodular tumors (≤ 3 nodules, each ≤ 3 cm) (Fig. 82-4). These so-called Milan criteria have been validated over the years, and a meta-analysis reported 5- and 10-year survival rates of ~70 and ~50%, respectively, similar to outcomes achieved in non-HCC transplantation indications. Perioperative mortality rates have been reduced to <3%. Transplantation simultaneously cures the tumor and the underlying cirrhosis, and it is associated with a low risk of recurrence, around 10–15% at 5 years. No immunosuppressive regimens or antitumor therapies after transplantation have demonstrated any preventive effect on recurrence. Milan criteria are integrated in the treatment strategy (BCLC 0 and A) and have also been adopted by the United Network for Organ Sharing (UNOS) pretransplant staging for organ allocation in the United States (stage T2). Aside from size and number, conventional contraindications for organ transplantation procedures (e.g., ABO incompatibility, comorbidities) are applied in this setting.

Liver transplantation has a couple of factors, such as cost and donor availability, that limit this procedure to <5% of HCC cases. The scarcity of donors represents a major drawback of liver transplantation. Donor scarcity varies geographically, and deceased liver donation is almost zero in some Asian countries. Due to the shortage of donors, median waiting times in Western programs is ~6–12 months, leading to 20% of candidates dropping off the list due to tumor progression before receiving the procedure. Predictors of dropout are neoadjuvant treatment failure, baseline AFP >400 ng/mL, and steady increase of AFP level >15 ng/mL per month. Several strategies have been proposed to overcome this limitation. First, apply neoadjuvant therapies in patients on the waiting list. Neoadjuvant treatments such as TACE or RF ablation have been assessed in the setting of cohort and cost-effectiveness studies. In principle, the use of these therapies is recommended when the waiting time exceeds 6 months, even though impact on long-term outcome is uncertain. Second, a priority policy has been established for patients enlisted. UNOS has implemented a scoring system based on the dropout risk.

The Milan criteria are universally used as the basis for transplant eligibility, and adherence to these criteria yields good posttransplant survival. Modest expansion of Milan criteria applying the “up-to-seven” criterion (i.e., those HCCs having the number 7 as the sum of the size of the largest tumor and the number of tumors) in patients without microvascular invasion achieves competitive outcomes. These pathologically defined criteria are being used in clinical practice to predict the expected outcome after transplantation. Similarly, *downstaging to Milan criteria* is currently defined as the reduction of HCC burden by locoregional treatments to achieve Milan staging before transplantation. This strategy leads to long-term 10-year survival rates of ~50%. Since policies for enhancing organ donation have reached a ceiling during the past several years, alternatives to donation have emerged. Living donor liver transplantation represents a plausible alternative that accounts of ~5% of total transplants performed globally.

Outcomes reported are similar to those with deceased liver donors, and it is recommended as an alternative option in patients on a waiting list exceeding 6 months. The risks and benefits of this procedure should take into account both donor (death is estimated in 0.3%) and recipient, a concept known as *double equipoise*. Due to the complexity of this treatment, it must be restricted to centers of excellence in hepatobiliary surgery and transplantation.

■ LOCOREGIONAL THERAPIES

LOCAL ABLATION RF ablation is recommended as the primary ablative technique (Fig. 82-4). The energy generated by RF ablation (heating of tissue at 80–100°C) induces coagulative necrosis of the tumor, producing a *safety ring* in the peritumoral tissue, which might eliminate small undetected satellites. Treatment consists of one or two sessions performed using a percutaneous approach, although in some instances, ablation with laparoscopy is needed. RF ablation is more effective in response rate and time to recurrence compared with the once-conventional percutaneous ethanol injection. HCC patients treated by RF ablation have 5-year survival rates of ~60% (Table 82-2). In tumors <2 cm, RF ablation achieves complete responses in >90% of cases with good long-term outcome and is competitive with resection in cost-effectiveness as first-line option. For BCLC A cases, RF ablation is the first-line treatment for single tumors 2–5 cm or up to three nodules, each ≤ 3 cm in diameter, unsuitable for surgery.

The failure rate of RF ablation increases in tumors >3 cm because of the heat loss due to perfusion-mediated tissue cooling within the area ablated. In tumors 3–5 cm in diameter, complete pathologic tumor necrosis of <50% has been reported. In particular, ~10–15% of tumors with difficult-to-treat locations, such as a subcapsular location or adjacent to the gallbladder, have a higher risk of incomplete ablation or major complications and can be approached by ethanol injection. Several approaches have been proposed to enhance the antitumor activity of RF ablation. Microwave ablation is the most widely used local image-guided technique alternative to RF. Theoretically, it provides major efficacy but higher complication rates in tumors >3 cm. Randomized trials comparing both techniques are needed. Other treatments, such as high-intensity focused ultrasound or stereotactic body radiotherapy for small tumors, have been studied in early clinical trials and are under investigation.

Chemoembolization TACE is the most widely used primary treatment for unresectable HCC worldwide and the first-line indication for patients with intermediate BCLC B stage (Fig. 82-4). Conventional chemoembolization (c-TACE) consists of the local hepatic artery administration of chemotherapy (either doxorubicin 50 mg/m² or cisplatin) mixed with an emulsion of lipiodol followed by obstruction of the feeding artery with sponge particles. c-TACE mainly benefits patients with liver-only disease, Child-Pugh A class or B class without ascites, good performance status (ECOG 0), and absence of branch or trunk vascular invasion. Median survival is ~20 months (compared to 16 months for pooled control arms). The best randomized phase 3 investigations have provided median survivals for TACE of 20–30 months in properly selected populations. Median objective response rates are 50–70%. In randomized studies, the treatment is either performed at a regular schedule of 0, 2, and 6 months (median number of sessions: 3) or on demand according to tumor response. TACE procedures should be stopped upon tumor progression or any other contraindication. Exceptionally, occurrence of a new small untreated nodule as the only progression feature can be considered for treatment. Around 50% of patients present with a limited postembolization syndrome of fever and abdominal pain related to ischemic injury and release of cytokines. Less than 5% of patients present with major complications (liver abscess, ischemic cholecystitis, or liver failure), and in <2% of cases, treatment-related death occurs.

Applicability of c-TACE in patients at intermediate stage is limited to half of cases, mostly as a result of the presence of liver failure (Child B or ascites or encephalopathy), technical contraindications to the procedure (i.e., impaired portal vein blood flow), or infiltrative/massive tumor burden (i.e., generally main tumor size >10 cm). Super-selective

TACE minimize the ischemic insult to nontumor tissue. According to guidelines, treatment-stage migration allows performing TACE on patients at early stages not suitable for surgical or ablative therapies. In selective studies, median survival times of 5 years have been reported in patients with single HCC treated by super-selective TACE. On the other hand, TACE performed beyond guidelines as a conventional practice in patients with advanced HCC yields poor outcomes.

Drug-eluting bead chemoembolization (DEB-TACE) differs from c-TACE in the use of more standardized embolic spheres of regular size embedded with chemotherapy. This strategy ensures drug release over a 1-week period, resulting in an enhancement of drug concentration within the tumor. DEB-TACE achieves similar antitumor activity (objective responses of ~60%) as c-TACE and is associated with significantly less systemic cytotoxic effects and better tolerance, but with no clear differences in clinical outcomes. Phase 2 and 3 studies have compared DEB-TACE with the combination of DEB-TACE plus sorafenib, orantinib, or brivanib, which are VEGF receptor inhibitors. Median survival in both arms of these international trials was 25–30 months.

Radioembolization and Other Intraarterial Therapies

Radioembolization using beads coated with yttrium-90 (Y-90)—an isotope that emits short-range β radiation—is the most promising alternative to TACE. Several phase II studies reported objective responses and overall outcome with a safe profile. Due to the lack of phase III trials, this treatment is currently not recommended in guidelines. Radioembolization requires prevention of severe lung shunting and intestinal radiation before the procedure. Around 20% of patients present with liver-related toxicity and 3% experience treatment-related death. Due to the minimally embolic effect of Y-90 microspheres, treatment can be

safely used in patients with portal vein thrombosis, a setting where survival results in phase II were encouraging. However, three randomised controlled trials, including two head-to-head trials against sorafenib and one trial combining radioembolization plus sorafenib versus sorafenib alone, did not show OS endpoint superiority. Thus, these treatments are not indicated in the advanced-stage scenario.

TACE should be distinguished from other intraarterial therapies, such as chemo-lipiodolization, which involves the delivery of an emulsion of chemotherapy mixed with lipiodol; bland transcatheter embolization, where no chemotherapeutic agent is delivered; and intraarterial chemotherapy, where no embolization is performed. None of these approaches is recommended due to the lack of survival benefit.

■ SYSTEMIC THERAPIES

Conventional systemic chemotherapy and radiotherapy have not produced survival advantages. Randomized studies also failed to show benefit with antiestrogen therapies and vitamin D derivatives. External-beam liver-directed radiotherapy (stereotactic body radiotherapy) efficacy is currently being tested with and without sorafenib in phase III trials. In 2007, a phase III trial demonstrated survival benefits for patients with advanced-stage disease treated with sorafenib, and lenvatinib showed similar effects to sorafenib in first-line treatment. Recently, the combination of atezolizumab with bevacizumab demonstrated survival benefits in the advanced setting when compared to sorafenib and has now become the standard first-line treatment. Three additional therapies, regorafenib, cabozantinib, and ramucirumab (only in patients with AFP >400 ng/mL), have been shown to benefit patients progressing on sorafenib (Fig. 82-5).

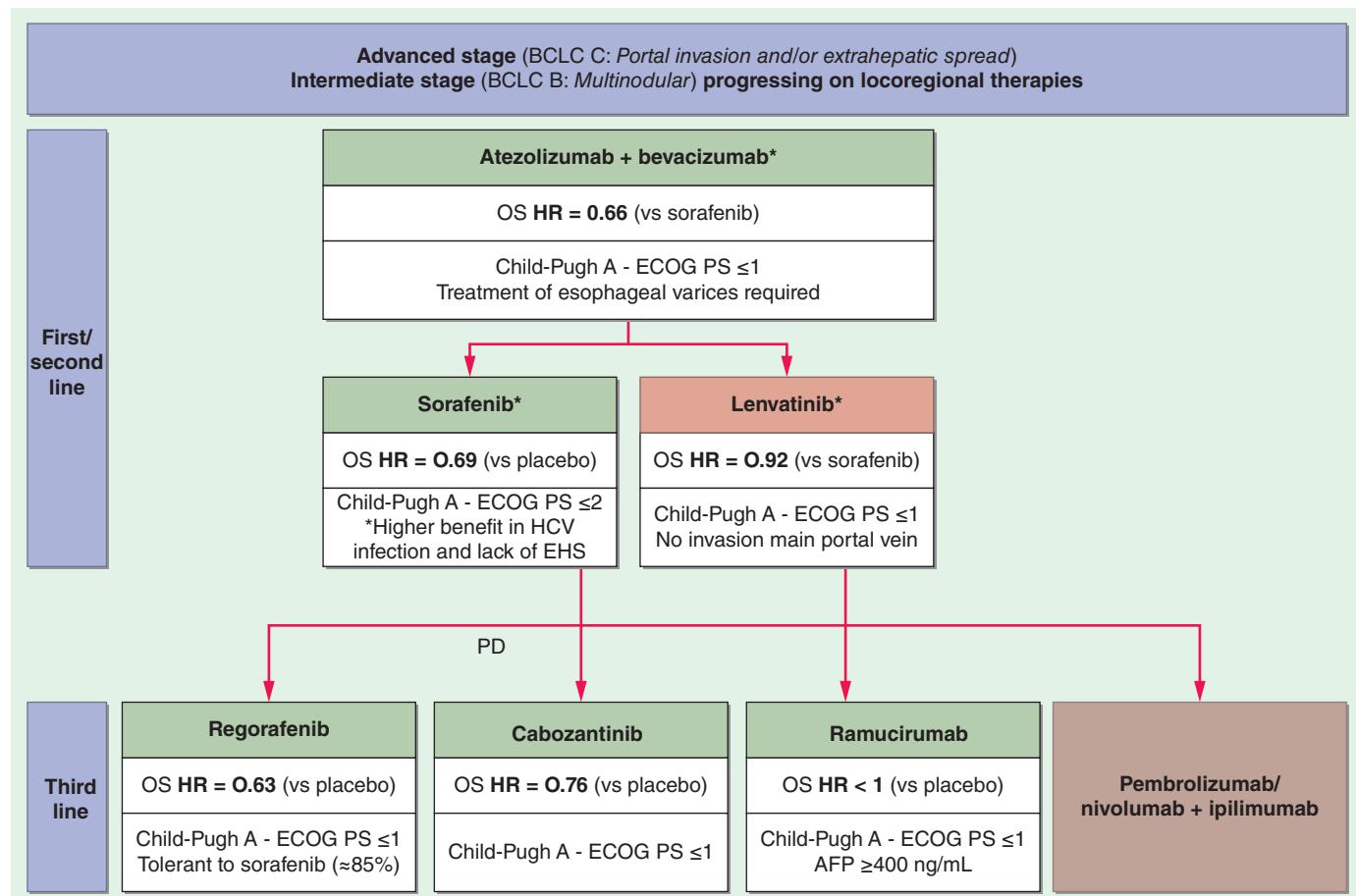


FIGURE 82-5 Treatment strategy for advanced hepatocellular carcinoma with systemic therapies. Drugs in green have positive results from phase 3 trials with a superiority design (atezolizumab plus bevacizumab, sorafenib, regorafenib, cabozantinib, and ramucirumab). Drugs in orange have positive results from phase 3 trials with a noninferiority design (lenvatinib vs sorafenib). Drugs in red have received accelerated approval from the U.S. Food and Drug Administration (FDA) based on promising efficacy results in phase 2 trials in the second-line setting (nivolumab, pembrolizumab, and nivolumab ipilimumab). Key details of the patient populations are provided. *Around 20% of patients can receive sorafenib or lenvatinib in first line due to contraindications to atezolizumab + bevacizumab. AFP, α -fetoprotein; BCLC, Barcelona Clinic Liver Cancer (classification); ECOG PS, Eastern Cooperative Oncology Group performance status; EHS, extrahepatic spread; HCV, hepatitis C virus; HR, hazard ratio; mRECIST, modified Response Evaluation Criteria in Solid Tumors; OS, overall survival. (Reproduced with permission from JM Llovet: Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat Rev Clin Oncol* 15:599, 2018.)

Molecular Targeted Therapies Atezolizumab (anti-PD-L1 checkpoint inhibitor) plus bevacizumab (antibody against VEGFA) has become the standard of care in first-line treatment for advanced HCC as a result of a positive phase 3 trial indicating superiority versus sorafenib in terms of survival (Fig. 82-5). Median survival with the combination was 19.2 months compared with 13.4 months for sorafenib. Combination treatment also improved progression-free survival and patient-reported quality of life outcomes. Objective response to the combination was 35.4% versus 13.9% for sorafenib. Adverse events also favored the combination (grade 3–4 adverse events, 36% vs 50% for sorafenib). The most common side effects associated with the combination were hypertension, proteinuria, and low-grade diarrhea, whereas autoimmune events were infrequent. Treatment-related adverse events leading to discontinuation of these two drugs was 15%. Upper gastrointestinal endoscopies are required before initiating the combination therapy for detection and treatment of varices to mitigate the risk of bleeding associated with bevacizumab. Thus, screening for varices is becoming standard before first-line therapy in HCC management.

Alternatively, sorafenib and lenvatinib are indicated for HCC in patients with well-preserved liver function (Child-Pugh class A) and with advanced tumors either as first-line treatment in patients with contraindications or with progression to the combination therapy

(Fig. 82-5). A phase III study comparing sorafenib versus placebo showed increased survival from 7.9 months to 10.7 months (hazard ratio [HR] 0.69; 31% reduction in risk of death). Patients with HCV-related HCC achieve significantly better outcomes with sorafenib, with a median survival of 14 months. No predictive biomarkers of responsiveness to sorafenib have been identified. The recommended daily dose of sorafenib is 800 mg. Median treatment duration is about 6 months. Treatment is associated with adverse events, such as diarrhea, hand-foot skin reactions, fatigue, and hypertension. These toxicities lead to treatment discontinuation in 15% of patients and dose reduction in up to half. This therapy cannot be administered to around one-third of the targeted patients due to primary intolerance, advanced age, or liver failure (ascites or encephalopathy). Active vascular disease, either coronary or peripheral, is considered a formal contraindication.

The efficacy of sorafenib probably results from a balance between targeting cancer cells and the microenvironment by blocking up to 40 kinases, including antiangiogenic (VEGF receptor [VEGFR], platelet-derived growth factor receptor [PDGFR]) and antiproliferative drivers (serine/threonine-protein kinase B-raf [BRAF] and mast/stem cell growth factor receptor [c-Kit]) (Fig. 82-6). Median time to progression on sorafenib is 4–5 months in phase III trials.

Another alternative to sorafenib is the multikinase inhibitor lenvatinib; it was noninferior in a phase 3 investigation (13.6 months vs

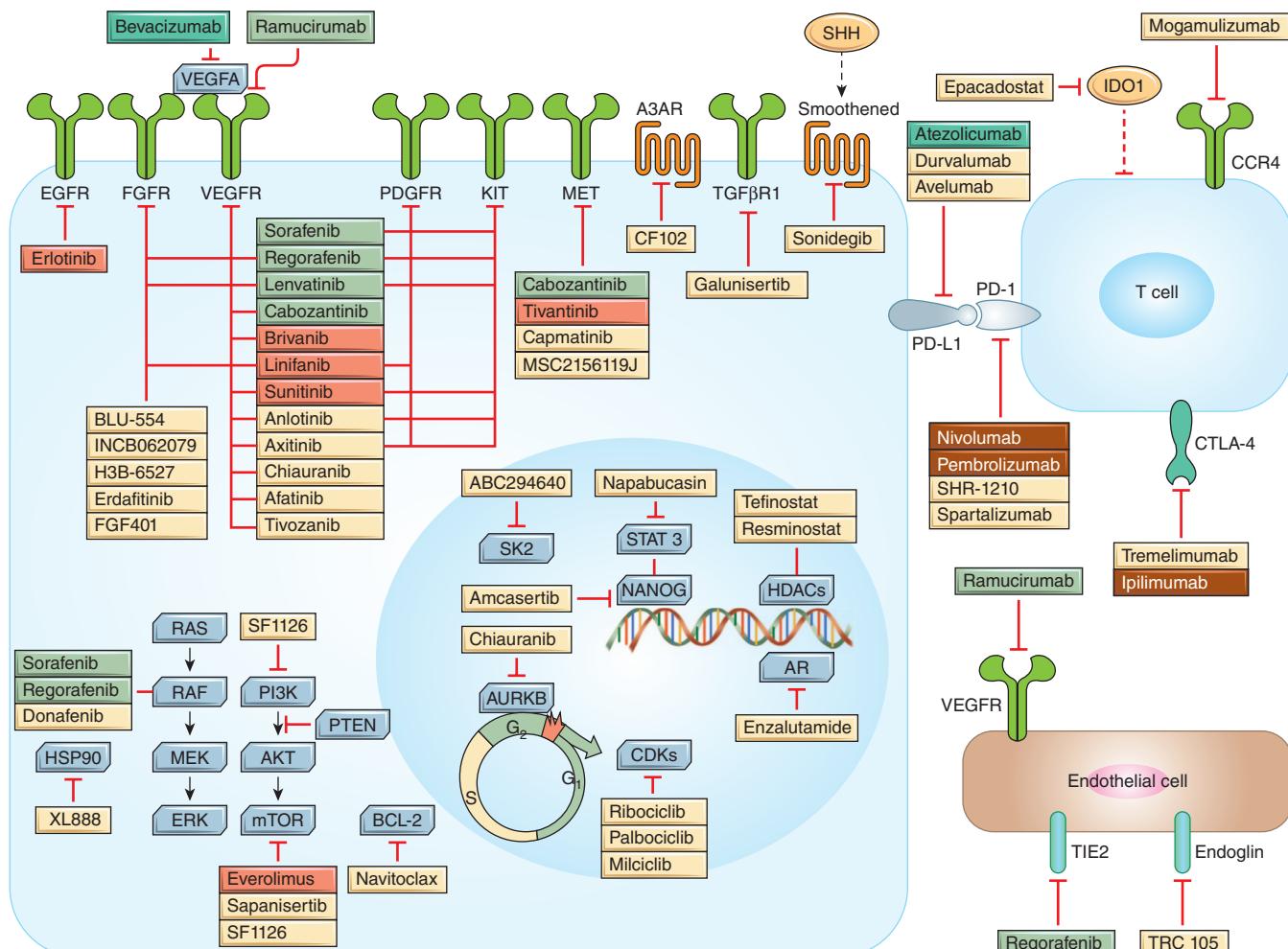


FIGURE 82-6 Molecularly targeted therapies for hepatocellular carcinoma and their target signaling pathways. Green boxes indicate drugs with positive results from phase 3 trials (atezolizumab plus bevacizumab, sorafenib, lenvatinib, regorafenib, lenvatinib, cabozantinib, and ramucirumab). Red boxes indicate drugs with negative results from phase 3 trials (everolimus, sunitinib, linifanib, erlotinib, brivanib, and tivantinib). Drugs in yellow boxes are currently in development for hepatocellular carcinoma in phase 1, 2, or 3 clinical trials. Brown boxes indicate drugs approved based on phase 2 trial data (pembrolizumab, nivolumab + ipilimumab). Dashed arrows and lines indicate indirect activities. A3AR, adenosine receptor A3; AR, androgen receptor; AURKB, aurora kinase B; BCL-2, apoptosis regulator BCL-2; CCR4, CC-chemokine receptor 4; CDKs, cyclin-dependent kinases; CTLA-4, cytotoxic T lymphocyte protein 4; HDAC, histone deacetylase; HSP90, heat shock protein 90; IDO1, indoleamine 2,3-dioxygenase 1; NANOG, homeobox protein NANOG; PD-1, programmed cell death protein 1; PD-L1, programmed cell death ligand 1; SHH, Sonic hedgehog; STAT3, signal transducer and activator of transcription 3; TIE2, angiopoietin 1 receptor. (Reproduced with permission from JM Llovet: Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol 15:599, 2018.)

12.3 months; HR 0.92) (Fig. 82-5). Lenvatinib induces objective responses in 24% of cases. The main side effects are hypertension, proteinuria, asthenia, diarrhea, and weight loss. This treatment induced grade 3–4 drug-related adverse events in 55% of patients, resulting in a withdrawal rate of ~15%.

Three drugs (regorafenib, cabozantinib, and ramucirumab) have shown survival benefits versus placebo in patients progressing on sorafenib, and two additional immune-based treatments have been approved by the U.S. Food and Drug Administration (FDA) based on promising phase 2 data (pembrolizumab and nivolumab plus ipilimumab) (Fig. 82-5). The median survival of patients progressing on first-line treatment is 8 months (obtained from patients allocated to the placebo arm).

A phase III study comparing regorafenib (a more potent multikinase inhibitor than sorafenib targeting similar kinases) versus placebo in patients progressing on sorafenib reported a benefit in survival from 7.8 to 10.6 months (HR 0.62; 38% reduction in risk of death) (Fig. 82-5). Response rate was 10%. Median time on treatment was 3.5 months. Prevalence of toxicity (hand-foot reaction, fatigue, and hypertension) was higher compared with reported toxicity from sorafenib, but adverse events only led to treatment discontinuation in 10% of cases. Cabozantinib, a multikinase VEGFR inhibitor with activity against both AXL and c-MET (Fig. 82-6), improves survival compared to placebo after progression on sorafenib (10.2 months for cabozantinib vs 8.0 months in the placebo arm; HR 0.76). The most common grade 3–4 adverse events were palmar-plantar erythrodysesthesia, hypertension, increased aspartate aminotransferase level, fatigue, and diarrhea. Ramucirumab, an anti-VEGFR-2 monoclonal antibody, is the only biomarker-guided therapy in HCC based on AFP levels. The randomized, placebo-controlled, phase 3 REACH-2 study selected patients with advanced HCC in second line with baseline AFP ≥ 400 ng/dL. Median survival for patients treated with ramucirumab was 8.1 months, compared to 5 months for patients receiving placebo. The most common grade 3–4 adverse events were hypertension, hyponatremia, and increased aspartate aminotransferase. Patients progressing after second-line therapy and patients with BCLC D stage should receive best supportive palliative care, including management of pain, nutrition, and psychological support.

Immunotherapy and Combinations The combination of the anti-PD-L1 antibody atezolizumab with the VEGFA inhibitor bevacizumab is the first regimen to improve survival in the first-line setting compared to sorafenib. In addition, two additional treatment regimens involving immunotherapies have been approved by the FDA as second-line therapies based on phase 2 data. Single-agent checkpoint inhibitor treatments, such as nivolumab and pembrolizumab, are associated with objective responses of 15–20%, which are durable in time, generally beyond 12 months. Less than 30% of patients experience grade 3–4 treatment-related adverse events. Neither regimen hit the primary endpoint of improved survival in phase 3 investigations compared with sorafenib (nivolumab) or placebo (pembrolizumab). The median survival for nivolumab of 16.4 months in first-line treatment was not superior to the survival time of 14.7 months for sorafenib. Similarly, in the second-line setting, the median survival for pembrolizumab of 13.9 months was not superior to the median survival of 10.6 months for placebo. Emerging regimens have shown signals of efficacy, such as lenvatinib plus pembrolizumab in first-line patients with advanced HCC and the combination of an anti-CTLA-4 (ipilimumab) and anti-PD-1 (nivolumab) in second-line patients.

CHOLANGIOPAPILLARY CARCINOMA

Cholangiocarcinoma (CCA) is classified according to its anatomic location as intrahepatic (iCCA; ~20–30%), perihilar (pCCA; ~50–60%), and distal (dCCA; ~20–30%). The latter two are also known as extrahepatic cholangiocarcinomas (eCCAs), with the second-order bile ducts acting as the separation point (Fig. 82-7). This classification is endorsed by the eighth edition of the *American Joint Committee on Cancer (AJCC) Staging Manual*. In addition, iCCA has been recognized as a distinct entity with specific ad hoc clinical practice guidelines. Treatment options beyond surgery are limited, and few molecular targeted therapies have been approved for its treatment. The three subtypes of CCA differ in their anatomic location, epidemiology and risk factors, cell of origin, pathogenesis, and treatment. iCCA originates from adult cholangiocytes, trans-differentiation of adult hepatocytes, and hepatic progenitor cells (cholangiocyte precursors) (Fig. 82-8), as opposed to HCC, which originates only from hepatic progenitor cells or adult hepatocytes. Mixed HCC-iCCA originates from hepatic progenitor cells, whereas eCCA arises from the biliary epithelium and peribiliary glands. Moreover, their mutational profiles also differ. FGFR2 fusions and IDH1/2 mutations mostly occur in iCCA, whereas ERBB2/3 amplifications and SMAD4 aberrations are characteristic of eCCA. Thus, clinical management and trials testing molecular therapies should be tailored according to each biological/anatomical subtype of CCA, as opposed to a common approach for all biliary tract cancers.

■ EPIDEMIOLOGY, RISK FACTORS, AND MOLECULAR TRAITS

CCA is the second most common liver cancer after HCC, with a 5-year survival of 10%. iCCA has globally increasing incidence and mortality rates. The incidence of iCCA varies according to exposure to risk factors, ranging from 1–2 cases per 100,000 inhabitants in Europe and North America to the highest incidence in some areas of Southeast Asia, particularly in Thailand (>80 cases per 100,000 inhabitants). The male-to-female ratio is 1.2. Overall, most cases occur with unknown risk factors. The classical risk factors for CCA development include primary sclerosing cholangitis (PSC), biliary duct cysts, hepatolithiasis, and Caroli's disease (congenital cystic dilation of the intrahepatic biliary tree). Parasitic biliary infestation with flukes (i.e., most common is *Opisthorchis viverrini* and *Clonorchis sinensis*) is a prevalent etiology in Asia that can be prevented with the antihelminth therapy praziquantel. PSC is a clear risk factor for iCCA and pCCA development, with a

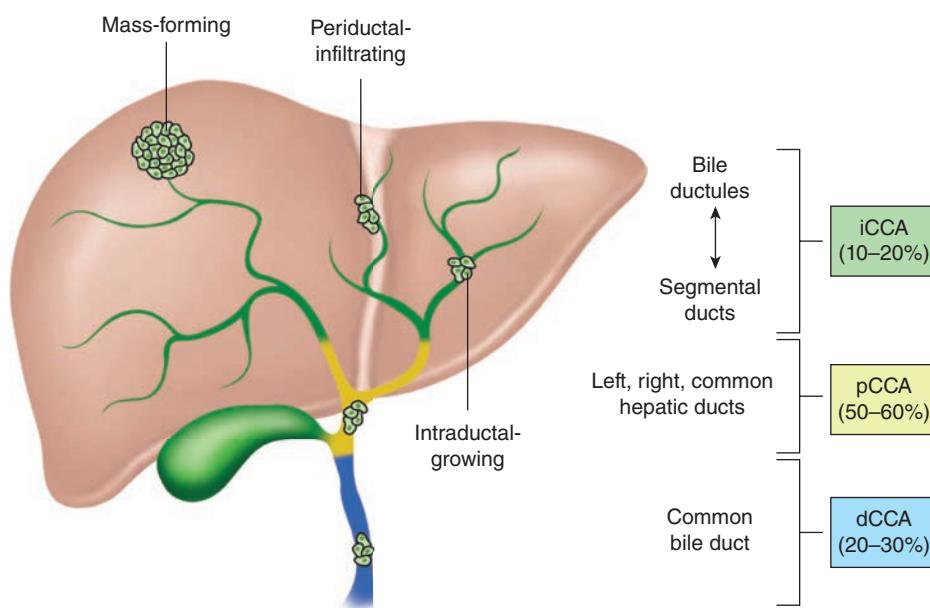


FIGURE 82-7 Anatomical classification of cholangiocarcinoma. Cholangiocarcinoma (CCA) is classified as intrahepatic (iCCA) and extrahepatic (eCCA). eCCA can be subclassified as perihilar (pCCA) and distal (dCCA). (Reprinted with permission from JM Banales et al: Cholangiocarcinoma 2020: The next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol* 17:557, 2020.)

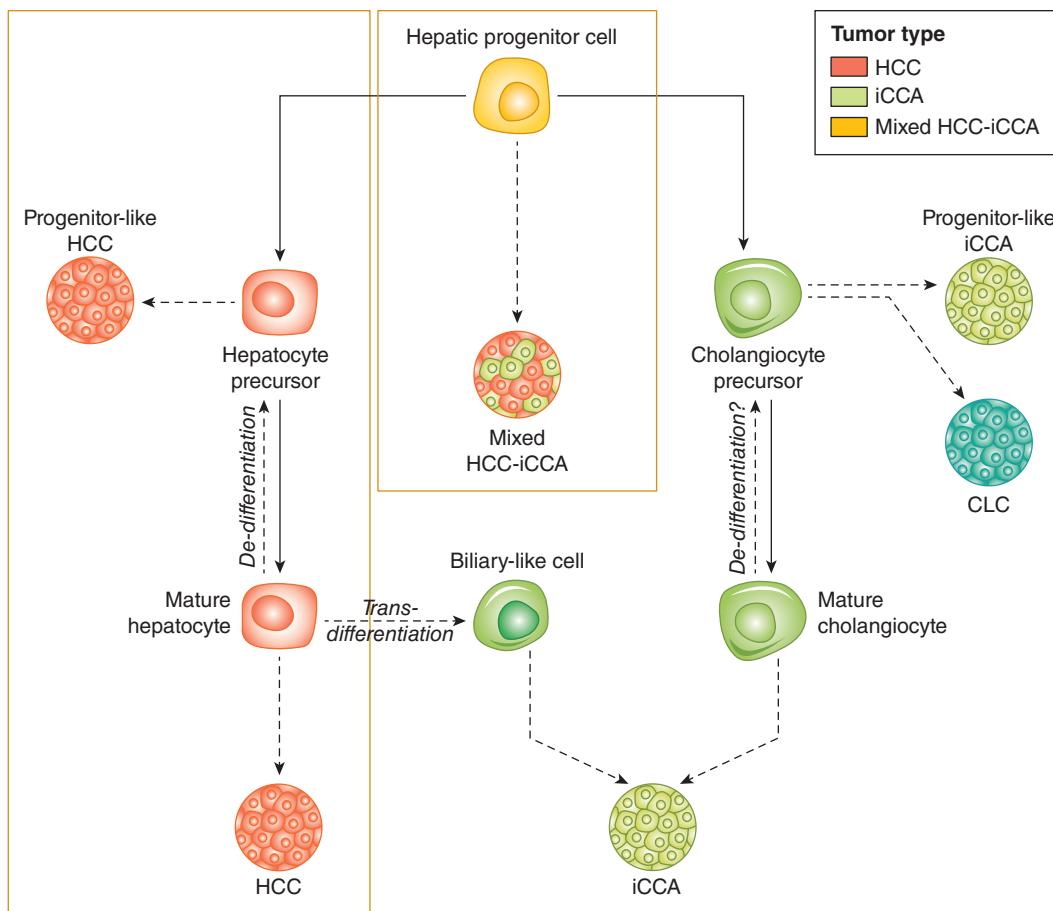


FIGURE 82-8 Cell of origin of liver cancer. Hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (iCCA) can develop from the neoplastic transformation of mature hepatocytes and cholangiocytes, respectively. There is evidence showing that hepatic progenitor cells (HPCs), their intermediate states, or de-differentiated hepatocytes can originate liver cancers with progenitor-like features, including mixed HCC-iCCA (e.g., cholangiocellular carcinoma [CLC]). Mature hepatocytes can be also reprogrammed into cells that closely resemble biliary epithelial cells and induce the onset of iCCA. (Printed with permission from ©Mount Sinai Health System.)

lifetime incidence ranging from 5 to 10%. Surveillance in PSC patients is recommended with annual imaging techniques and CA 19-9 serum determination. Common risk factors for HCC, such as HBV and HCV infection and cirrhosis, have been associated with iCCA development. Sweetened beverages were reported to constitute an additional risk factor in the development of eCCA and gallbladder carcinoma in a population cohort study.

Molecular Classification and Drivers No molecular classification of CCA has been established. Genomic studies have provided insight on two subclasses of iCCA, a proliferation subclass—characterized by activation of oncogenic signaling pathways (including RAS and MET)—and an inflammation subclass, characterized by activation of inflammatory pathways, overexpression of cytokines, and STAT3 activation. Similarly, a molecular classification of eCCA has been proposed, dividing tumors into four categories (metabolic, proliferation, mesenchymal, and immune) based on molecular traits. The hypothesis that the proliferation class with enrichment of *ERBB2/3* mutations might respond to monoclonal antibodies against this receptor and the immune class might respond to checkpoint inhibitors has not been tested or confirmed. The iCCA mutation portrait is characterized by ~50–60% of tumors having at least one targetable driver including *FGFR2* fusion events (~25%); mutations in *IDH1/2* (15%), *KRAS* (15%), *BRAF* (5%), and *EGFR* (3%); and amplifications in *FGF19/CCDN1* (4%). Although mutations in *TP53* (~30%) and *KRAS* (~25%) are more common in eCCA than in iCCA, some molecular drivers are specific for subtypes, such as fusion of *PRKACA* or *PRKACB* for eCCA or *ERBB2* amplifications (~20%) for gallbladder cancer. Liver fluke-associated CCAs have a higher incidence of *TP53* and *SMAD4* mutations. Host genetic polymorphisms predisposing to CCA have not been established.

■ INTRAHEPATIC CHOLANGIOPANCREATOGRAPHY

Diagnosis Diagnosis of iCCA requires pathologic confirmation. Guidelines currently do not recommend surveillance for early diagnosis because at-risk populations are ill-defined. Cirrhotic patients at risk of HCC development are enrolled in surveillance programs and can benefit from early detection of iCCA. Otherwise, incidental diagnosis occurs due to cross-sectional imaging performed for other reasons. In most cases, iCCA is diagnosed at advanced stages where symptoms such as weight loss, malaise, abdominal discomfort, or jaundice are present. Pathologic diagnosis of iCCA is based on the World Health Organization (WHO) criteria. Differential diagnosis should be established with metastatic adenocarcinoma and mixed iCCA-HCC tumors, which may require evaluation of markers such as Hep-Par-1, GPC3, HSP70, and glutamine synthetase markers. Imaging studies with CT/MRI are not accurate enough to establish iCCA noninvasive diagnosis. Dynamic CT scanning characterizes 80% of iCCAs as liver mass-forming tumors with progressive contrast uptake from the arterial to the venous/delayed phase. MRI dynamic images also show peripheral enhancement in the arterial phase followed by progressive filling in of the tumor. Atypical radiologic behavior with arterial enhancement recapitulating HCC occurs in 10% of cases. MRI with cholangiopancreatography is useful to visualize the ductal system and vascular structures. Guidelines do not recommend PET scan for diagnosis. Tumor biomarker CA 19-9 at a cutoff level of 100 U/mL has prognostic significance but lacks accuracy (sensitivity and specificity of ~60%) for early diagnosis.

Radiologic criteria are inadequate for iCCA diagnosis in cirrhotic patients. However, in noncirrhotic patients, guidelines endorse a presumed diagnosis of iCCA (i.e., venous phase contrast enhancement on dynamic CT/MRI) if resection is considered. Assessment of disease

extent (venous or arterial invasion and extrahepatic disease) and resectability is best accomplished with CT and/or MRI studies. Doppler ultrasound is accurate in defining vascular invasion. Before surgery, PET scanning may be considered to rule out an occult primary or metastatic site.

Staging System The staging system for iCCA resected cases is based on the TNM staging as per the eighth edition of the AJCC/International Union Against Cancer (UICC) staging. T1 tumors are solitary without vascular invasion and can be divided into T1a or T1b if tumor size is ≤ 5 cm or > 5 cm, respectively; T2 disease includes multiple tumors (e.g., multifocal disease, satellitosis, intrahepatic metastasis) or presence of vascular invasion (microvascular or major vascular invasion); T3 tumors perforate the visceral peritoneum; and T4 disease includes tumors involving local extrahepatic structures by direct invasion. Regional lymph node metastasis in the hilar, periduodenal, and peripancreatic nodes is considered N1 disease, while distant spread is considered M1 disease. TNM stages IA, IB, II, and IIIA overlap with T status, whereas stage IIIB includes T4N0 or N1M0 disease and stage IV includes M1 disease.

TREATMENT

After adopting the TNM staging system, the International Liver Cancer Association (ILCA) guidelines for management of iCCA proposed a treatment algorithm (Fig. 82-9), adapted and updated with the new treatment modalities accepted. Overall, most of the treatments endorsed have a modest level of evidence. Surgical resection represents the sole curative treatment option in 30–40% of patients, with

a median survival of 51 months in properly selected candidates. In noncirrhotic individuals, the best candidates for resection are patients at TNM stage I-II, whereas in patients with cirrhosis, liver function should be assessed as previously described for HCC. Preoperative disease assessment should discard vascular invasion, N1, and M1. Lymphadenectomy of regional nodes is recommended given its prognostic value. The main predictors of recurrence (~50–60% at 3 years) and survival are identified at the pathologic examination, including presence of vascular invasion, lymph node metastases, and poor differentiation. A phase 3 trial (BILCAP trial) including all types of CCA in a prespecified per-protocol analysis reported improved survival with adjuvant therapy (53 months vs 36 months; adjusted HR 0.75). Based on this trial, American Society of Clinical Oncology guidelines recommend adjuvant capecitabine for a period of 6 months. Other adjuvant regimens, such as gemcitabine monotherapy or a combination of gemcitabine and oxaliplatin, did not improve OS. Liver transplantation remains controversial, and few studies have reported good outcomes for single tumors ≤ 2 cm.

Nonsurgical candidates have a dismal life expectancy. Overall, patients at stage III might be considered for locoregional therapies, such as chemoembolization or radioembolization, but the level of evidence is low and is mostly based on cohort studies. A meta-analysis of 14 trials testing locoregional therapies reported median survival times of 15 months. External-beam radiation therapy is not recommended as standard therapy. At more advanced stages (stage IV) in patients with an ECOG of 0–1, systemic chemotherapy with the combination of gemcitabine and cisplatin is considered the standard of practice, yielding median survival times of 11.7 months compared to 8 months

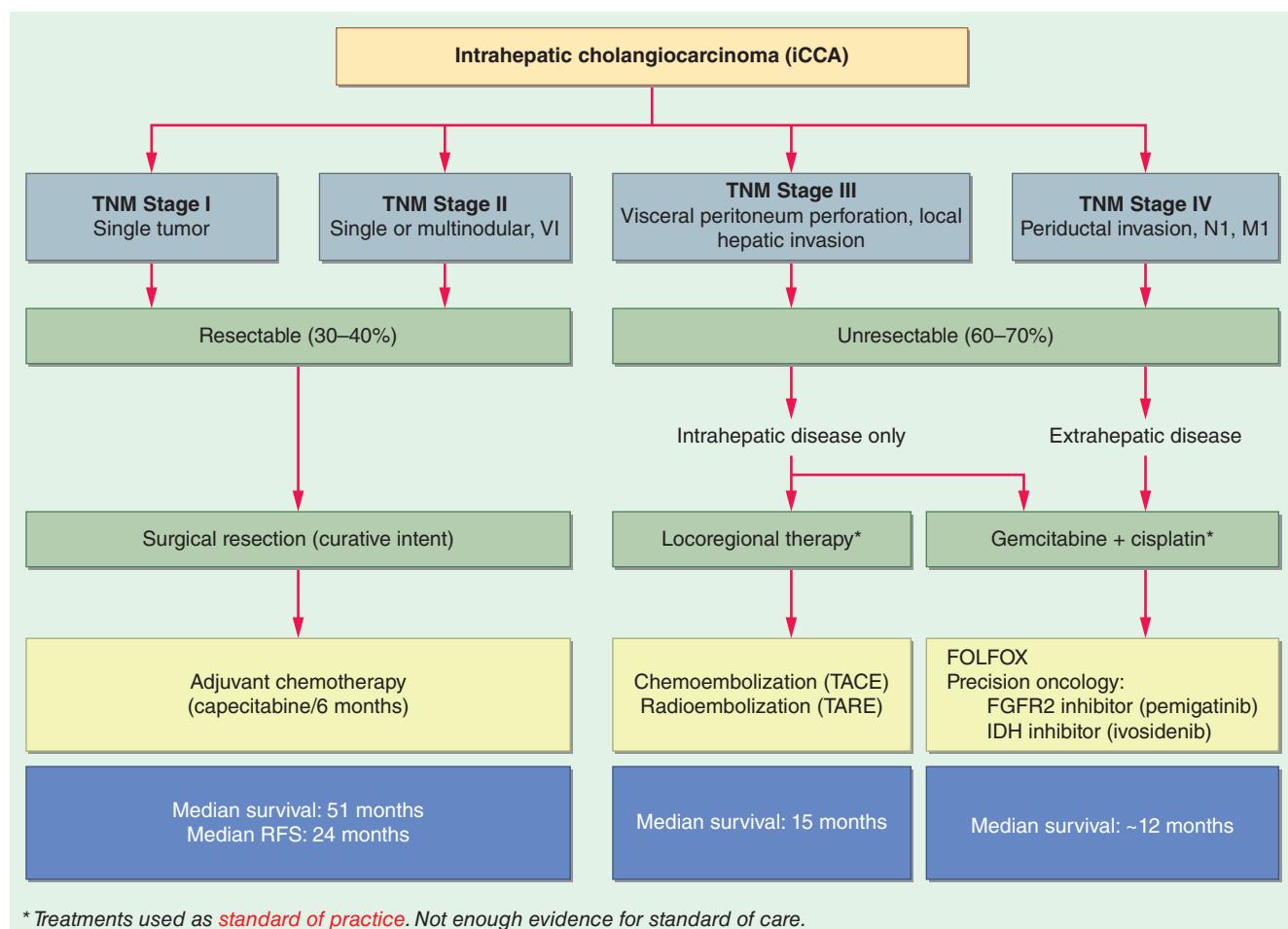


FIGURE 82-9 Staging and treatment schedule for intrahepatic cholangiocarcinoma (iCCA) proposed by the International Liver Cancer Association. FOLFOX, leucovorin, fluorouracil, and oxaliplatin; RFS, recurrence-free survival; TACE, transcatheter arterial chemoembolization; TARE, transarterial radioembolization; TNM, tumor-node-metastasis. (Reproduced with permission from J Bridgewater et al: Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol* 60:1268, 2014.)

for gemcitabine alone. This recommendation for first-line treatment of advanced tumors is based on a subgroup analysis of 80 iCCA patients included in a large randomized phase III trial ($n = 410$, ABC-02 Trial) of patients with advanced biliary tract tumors. In the second-line setting, a phase 3 study randomized patients who had progressed on cisplatin and gemcitabine to mFOLFOX (leucovorin, fluorouracil, and oxaliplatin) versus best supportive care. The chemotherapy regimen showed an improvement in median OS to 6.2 months (adjusted HR 0.69).

Two molecular targeted therapies have been approved in the second-line setting in iCCA patients with *IDH1/2* mutations or *FGFR2* aberrations. A phase 3 trial compared ivosidenib, an IDH-1 inhibitor, versus placebo; ivosidenib improved progression-free survival (2.7 vs 1.4 months; HR 0.37) and OS. A single-arm phase 2 study assessing pemigatinib (*FGFR2* inhibitor) in iCCA patients with *FGFR2* fusions showed a median survival of 21 months and an objective response rate of 35%.

Mixed HCC-iCCA is a rare neoplasm accounting for <0.5% of all primary liver cancers. Diagnosis is based on pathology. The 2010 WHO classification defined two subtypes: the classical and the stem cell feature type. Molecular data have defined a third unique entity, cholangiocellular carcinoma, with distinct molecular traits and better outcome. Due to its low incidence, the demographic features and clinical behavior of these tumors remain ill-defined. Survival and management are similar to iCCA.

EXTRAHEPATIC CHOLANGIOPANCREATOGRAPHY

Perihilar and Distal Cholangiocarcinoma The eighth edition AJCC/UICC TNM staging classification has established pCCAs as tumors that arise between the second-order bile ducts up to the insertion of the cystic duct, whereas dCCAs arise from this point to the ampulla of Vater (Fig. 82-7). Thus, dCCA can be difficult to distinguish from early pancreatic cancer. Both entities have a similar diagnostic approach. Acute onset of painless jaundice occurs in 90% of patients with pCCA, and 10% present with cholangitis. Primary biliary cholangitis with a cutoff for CA 19-9 >129 U/mL is suspicious for CCA. Imaging assessment starts with CT and MRI; they have a good sensitivity and specificity (>85%) for detecting the degree of bile duct involvement and hepatic and portal vein invasion. MRI cholangiography is optimal for defining the extent of the bile duct lesion. Ruling out IgG4 cholangiopathy by assessing serum IgG4 is mandatory. As a second step, endoscopic retrograde cholangiography with brushing to explore cytology and fluorescence in situ hybridization (FISH) for exploring polysomy are recommended. FISH enhances the sensitivity of cytology from 20 to ~40%.

Diagnosis is based on pathology. The treatment algorithm for pCCA indicates that in cases of a dominant stricture with positive cytology/biopsy or polysomy, a lymph node biopsy via endoscopic ultrasound should be obtained. pCCA with negative lymph node involvement is best treated by surgery, resection, or transplantation, the sole curative options. Staging laparoscopy is recommended to exclude metastatic disease before surgery; metastases occur in 15% of cases. Resection entails hepatic and bile duct removal and Roux-en-Y-hepatojunostomy with regional lymphadenectomy. Bilobular involvement is considered a surgical contraindication. Perioperative mortality is as high as 10%, mostly as a result of liver failure. In a few referral centers, unresectable single pCCA <3 cm without dissemination can be considered for liver transplantation with neoadjuvant chemoradiation. This procedure is associated with 5-year survival rates of ~70%. If lymph node involvement is present, systemic chemotherapy can be considered along with biliary tract stenting. Surgical resection (Whipple procedure) is the primary option for management of dCCA, a procedure that achieves a median survival of 2 years and 5-year survival rates of ~25%. Main contraindications for resection are presence of distant lymph node involvement, metastases, or major vascular invasion. At the pathologic examination, perineural invasion, lymph node metastasis, R0 resection (absence of residual tumor at pathologic examination), and tumor differentiation are predictors of survival. Adjuvant therapy with capecitabine for 6 months is accepted based on the BILCAP study.

Consensus statements endorse first- and second-line chemotherapy strategies for unresectable eCCA similar as for iCCA. No molecular targeted therapies are available for these entities.

GALLBLADDER CANCER

Gallbladder cancer is the most common cancer of the biliary tract worldwide. The estimated cases of gallbladder cancer in the United States in 2020 were 11,980, more than CCA. The female-to-male ratio is 3:1. Cholelithiasis is the major risk factor, but <1% of patients with cholelithiasis develop this cancer. Gallbladder polyps at risk of transformation are those ≥10 mm in diameter. Early cases are discovered incidentally at routine cholecystectomy. Clinical symptoms, such as jaundice, pain, and weight loss, are associated with advanced stages. Staging of gallbladder cancer follows the TNM classification. The most accurate technique to define staging and vascular and biliary tract invasion is magnetic resonance cholangiopancreatography. CT and PET scans can be also useful for preoperative staging.

The mainstay of treatment is surgical, either simple or radical cholecystectomy (partial hepatectomy and regional lymph node dissection) for stage I or II disease, respectively. Only ~20% of patients are candidates for surgery with curative intent. Survival rates are near 80–90% at 5 years for stage I disease and range from 60 to 90% at 5 years for stage II disease. Regional nodal status and the depth of tumor invasion (T status) are the two most important prognostic factors. Adjuvant therapy with capecitabine is recommended in R0 cases. Gallbladder cancers at stage III and IV are considered unresectable. For patients with ECOG of 0–1, chemotherapy with gemcitabine and cisplatin is the standard of practice based on data from the subgroup analysis including 181 patients with gallbladder cancer in the setting of two clinical trials. Overall, median survival is 10–12 months in advanced cases. Percutaneous transhepatic drainage is indicated in case of biliary obstruction. Radiotherapy is not effective.

OTHER MALIGNANT LIVER TUMORS

FIBROLAMELLAR HEPATOCELLULAR CARCINOMA

Fibrolamellar hepatocellular carcinoma (FLC) is a rare form of primary liver cancer that typically affects children and young adults (10–30 years of age) without background liver disease. FLC accounts for 0.85% of all primary hepatic malignancies in the United States, and its incidence rate is 0.02 cases per 100,000 inhabitants. FLC is considered a unique entity with a specific fusion oncogene *PRKACA-DNAJB1* present in 80–100% of cases. A few mutations have been described, all at a level of <10%. FLC has a better prognosis than HCC, probably due to the absence of cirrhosis and the earlier age of presentation. Surgical resection is the mainstay of treatment, and indications are less restrictive than for HCC. A retrospective series of 575 FLC cases reported a median survival of 70 months after resection. At advanced stages, the expected outcome is <20 months. Chemotherapy is not effective, and there is no standard of care.

HEPATOBlastoma

Hepatoblastoma (HB) is the most frequent primary liver tumor in children. The incidence of the disease is 1.5 cases per 1,000,000. Background liver disease is rare in these patients. WNT signaling plays a major role, with *CTNNB1* mutations (70%) as the most frequently reported molecular event. Overexpression of IGF2 and genes in the 14q32 *DLK1/DIO3* locus are also prevalent. Resection followed by chemotherapy with doxorubicin is the mainstay treatment strategy. A study including 1605 patients randomized in eight clinical trials reported better outcome for patients with stage I-II of the PRETEXT (Pretreatment Extent of Tumor) classification (out of four stages), age <3 years, AFP >1000 ng/mL, and absence of metastases. As opposed to HCC, low AFP indicates poor prognosis. The best candidates (stage I or II with small tumors, age <3 years, and AFP >100 ng/mL) achieve 5-year disease-free survival after resection of 90%, compared with 5-year disease-free survival of 20–30% in the worst candidates (metastatic disease and AFP <100 ng/mL).

BENIGN LIVER TUMORS

The most common benign liver tumors are hemangiomas, focal nodular hyperplasia (FNH), and hepatocellular adenomas (HCA). Most benign tumors are identified incidentally by abdominal ultrasound or other imaging techniques. *Hemangiomas* are present in ~5% of the general population and are diagnosed by ultrasound, except in cirrhotic patients or oncology patients in whom contrast-enhanced imaging (contrast-enhanced ultrasound, CT, or MRI) is required. Conservative management is appropriate and follow-up is not recommended. Exceptionally, growing lesions causing symptoms by compression can be considered for resection. FNH is a benign tumor present in <2% of the population and occurs mostly in females aged 40–50 years. FNH is a polyclonal hepatocellular proliferation due to an arterial malformation. MRI has the highest diagnostic accuracy with a specificity of 100% when typical imaging features are present (homogeneous enhancement in the arterial phase with a central scar). Atypical FNH requires biopsy for diagnosis. Treatment is not recommended since these tumors do not degenerate or cause complications. In exceptional cases of expanding symptomatic lesions, surgery is the treatment of choice.

Hepatic adenomas are clonal benign proliferations resulting from single-gene driver mutations. HCAs have a low prevalence of 0.001% of the population and are frequently diagnosed in women aged 35–40 years. The female-to-male ratio is 10:1, and the main risk factors are oral contraceptives in females and use of anabolic androgenic steroids in male body builders. HCAs have the potential for hemorrhage and HCC development, particularly when >5 cm. Molecular classification of HCA is defined as follows: (1) HCA with *CTNNB1* mutations (10–20%) are at risk of HCC development and are present in men treated with androgens; (2) inflammatory adenomas (50–60%) are associated with single mutations (*Gp130*: 65%) and are more prevalent in females with obesity or diabetes; and (3) adenomas with inactivated *HNF1A*. Diagnosis is based on MRI, which correlates with molecular subtypes in 80% of cases (inflammatory and HNF-1A type). For defining HCA with *CTNNB1* mutations, biopsy is required. Upon diagnosis, discontinuation of oral contraceptives and weight loss are recommended. Resection is indicated in all cases of >5 cm, in men, or in those with *CTNNB1* mutation. For HCA <5 cm, 1-year follow-up is recommended. In case of active HCA bleeding, embolization followed by resection is the treatment of choice. The presence of multiple HCAs is common, and guidelines endorse treating them based on the size of the main nodule.

FURTHER READING

- BANALES J et al: Cholangiocarcinoma 2020: The next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol* 17:557, 2020.
- BRIDGEWATER J et al: Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol* 60:1268, 2014.
- EASL-EORTC CLINICAL PRACTICE GUIDELINES: Management of hepatocellular carcinoma. *J Hepatol* 69:182, 2018.
- FINN RS et al: Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *N Engl J Med* 382:1894, 2020.
- LLOVET JM et al: Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* 359:378, 2008.
- LLOVET JM et al: Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat Rev Clin Oncol* 15:599, 2018.
- LLOVET JM et al: Locoregional therapies in the era of molecular and immune treatments for hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* 18:293, 2021.
- LLOVET JM et al: Hepatocellular carcinoma. *Nat Rev Dis Primers* 7:6, 2021.
- MARRERO J et al: Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology* 68:723, 2018.
- MAZZAFERRO V et al: Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med* 334:693, 1996.
- RIZVI S et al: Cholangiocarcinoma: Evolving concepts and therapeutic strategies. *Nat Rev Clin Oncol* 15:95, 2018.

SCHULZE K et al: Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 47:505, 2015.

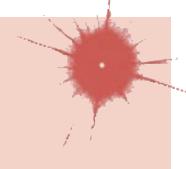
VILLANUEVA A: Hepatocellular carcinoma. *N Engl J Med* 380:1450, 2019.

ZUCMAN-RONSI J et al: The genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology* 149:1226, 2015.

83

Pancreatic Cancer

Daniel D. Von Hoff



Pancreatic cancer is the third leading cause of death from cancer in the United States, with >57,000 Americans diagnosed and >47,000 dying from the disease each year. Unfortunately, pancreatic cancer is projected to be the second leading cause of death from cancer in the United States by 2030. Worldwide, pancreatic cancer is the eleventh most common cancer with 458,000 new patients diagnosed and >432,000 deaths (seventh cause of cancer deaths). Pancreatic cancer currently has the worst survival rate of any cancer with an overall 5-year survival (regardless of stage) of ~8.2%. However, that situation is changing. In particular, (1) knowledge about specific molecular subsets of the disease has become crucial to provide the best possible care for patients, and (2) the application of treatment that improves survival for patients with advanced disease used either after surgery or even earlier in the disease has improved survival.

EPIDEMIOLOGY

Pancreatic cancer accounts for 3.2% of all new cancer cases in the United States and for 7.8% of all deaths from cancer in the United States. The lifetime risk of developing pancreatic cancer is ~1.7%. The incidence of pancreatic cancer has been increasing about 1.03% per year. Pancreatic cancer is more common with increasing age and more common in men than in women. The 5-year survival rate for all stages has only increased from 3% in 1975 to 9% in 2015. The latest information from the U.S. Surveillance, Epidemiology, and End Results (SEER) database predicts that the 5-year survival for patients with localized pancreatic cancer is about 37%, 12% for those with regional disease, and 3.1% for patients with advanced metastatic disease. Pancreatic cancer is more common in developed countries (although generally it tracks with the prevalence of smoking). The incidence is highest in Western Europe and North America followed by other areas in Europe, Australia, New Zealand, and South-Central Asia. The population at greatest risk are women living in Scandinavian countries, while the lowest risk is seen for women living in middle Africa.

RISK FACTORS

Age is one of the greatest risk factors for pancreatic cancer with median age at diagnosis of 70 years (the disease is most frequently diagnosed in the 65–79 age group; for men, 65–69; for women, 75–79). The number of new cases per 100,000 persons and the number of deaths per 100,000 persons are higher for males and for blacks of both sexes. Both the number of cases and the number of deaths per 100,000 people are lower for American Indian/Alaskan natives and Asian Pacific Islanders. Both the number of cases and deaths are intermediate for the Hispanic population. People who have a non-O blood type are at higher risk of developing pancreatic cancer.

Environment The greatest risk factor for pancreatic cancer is cigarette smoking. The risk correlates with the increased number of cigarettes smoked and persists for at least 10 years after smoking cessation. About 30% of pancreatic cancer is caused by smoking. Exposure to cadmium as part of cigarette smoking or via exposure to welding,

soldering, or dietary exposure has been weakly associated with an increased risk of pancreatic cancer.

Although dietary factors are often difficult to interpret, high intakes of fat or meat (particularly well-done barbequed meat) are risk factors. High intakes of citrus fruits and vegetables are associated with a decreased risk. Coffee and low-to-moderate alcohol consumption are not associated with an increased risk for pancreatic cancers, while consumption of sugary carbonated drinks has been associated with an increased risk.

Microbiome To date, no solid evidence links *Helicobacter pylori* infection and pancreatic cancer. Some data link the oral microbiome associated with poor dentition to pancreatic cancer, but the evidence is very thin.

Hereditary/Genetics Hereditary factors may account for 10–16% of all pancreatic cancers. Family members of patients with pancreatic cancer should seek participation in an early detection program with genetic counseling, definition of risk, and if appropriate, periodic MRI screening of the abdomen, although this recommendation is not yet based on research data. In addition, the identification of any pancreatic cancer-associated germline mutations could lead to specific and effective new therapeutics for patients with these abnormalities in their tumors. **Table 83-1** identifies the various germline mutations along with their familial cancer syndromes where an increased risk for pancreatic cancer is known.

Knowing the patient has a *BRCA2* or *PALB2* germline mutation or any of the above mutations should lead one to not only refer the patient's relatives to an early detection or high-risk individual clinic but also realize that for patients with a *BRCA2/PALB2* germline mutation consideration for treatment with a poly (ADP-ribose) polymerase (PARP) inhibitor should be considered (see below). Other germline mutations are under study to determine their increased risk of pancreatic cancer, including *CFTR*, *PRSS2*, *CDK4*, *FANCC*, *PALLD*, *APC*, *ATM*, *BMPR1A*, *BRCA1*, *EPCAM*, *MEN1*, *MLH1*, *MSH2*, *MSH6*, *NF1*, *PMS2*, *SMAD4*, *TP53*, *TSC1*, *TSC2*, and *VHL*. Some of these mutations are associated with pancreatic neuroendocrine tumors (**Chap. 84**).

In addition to the recognized genetic syndromes, other possible familial pancreatic cancer genes have not yet been discovered. For example, a family history of pancreatic cancer is associated with a 13-fold increase in the disease. If you have one first-degree relative, the risk is increased 4.6-fold, having two first-degree relatives increases the risk 6.4-fold, and three or more first-degree relatives confers a 32-fold increased risk. The risk is also increased if a relative developed pancreatic cancer at <55 years old.

TABLE 83-1 Germline Mutations, Their Familial Cancer Syndrome, and Fold Risk of Pancreatic Cancer

GERMLINE MUTATION	FAMILIAL CANCER SYNDROME	ESTIMATED INCREASED RISK (FOLD) OF PANCREATIC CANCER
<i>BRCA2</i> ^a	Familial breast/ovarian cancer	2–6
<i>PALB2</i> (partner and localizer of <i>BRCA2</i>)	Familial breast cancer and others	~sixfold
<i>p16/CDKN2A</i>	Familial atypical multiple mole melanoma (FAMMM)	15–18
<i>STK11 (LKB1)</i>	Peutz-Jeghers syndrome	76–140
<i>PRSS1</i> or <i>SPIN11</i> ^b	Hereditary (familial) pancreatitis	53
<i>ATM</i>	Ataxia-telangiectasia	Not yet established
<i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> , <i>PMS2</i>	Hereditary nonpolyposis colorectal syndrome or Lynch syndrome ^c	9–30

^aParticularly common in individuals with Ashkenazi Jewish heritage. ^bForty percent chance of pancreatic cancer by the age of 70. ^cVery important because this is associated with microsatellite instability, which is a marker for response to an anti-PD-1/PD-L1 agent.

Other Considerations Most patients with pancreatic cancer relate that they have had developing symptoms over the past few years. Thus, early detection of the disease is possible when the index of suspicion is high.

Medical Conditions Chronic pancreatitis that is nonfamilial is also associated with an increased risk of pancreatic cancer (2.3–16.5-fold increase). Risk is also increased in people with chronic pancreatitis associated with cystic fibrosis or tropical pancreatitis.

A clear association exists between diabetes mellitus (both type 1 and type 2) and pancreatic cancer. Whether this is a causal association or whether the diabetes is the result of the cancer is not exactly clear. What is clear is that when a person presents with new-onset type 2 diabetes, they should be considered at risk for having pancreatic cancer. The excessive insulin or insulin-like growth factors associated with adult-onset diabetes and metabolic syndrome may promote pancreatic carcinogenesis.

Obesity is considered a possible risk factor for pancreatic cancer. A high body mass index (BMI) ≥ 30 is associated with a doubling of the risk of pancreatic cancer. Since obesity is a risk factor for diabetes, the contribution of obesity alone is unclear. Interestingly, patients with severe obesity who undergo a gastric bypass experience a reduction in the incidence of gastrointestinal (GI) cancer, including pancreatic cancer, by >30% in the first 3 years (along with a dramatic decrease in their hemoglobin A_{1c} and blood glucose). Physical inactivity also has been associated with an increased risk in pancreatic cancer.

■ PATHOLOGY AND MOLECULAR CONSIDERATION

Location The posterior location of the pancreas in the abdomen is likely one of the issues that leads to a late diagnosis (**Fig. 83-1A**).

Pathology Cancers of the pancreas can be divided into neoplasms of the endocrine pancreas (**Chap. 84**) and tumors of the exocrine pancreas. The most common neoplasm of the exocrine pancreas and most deadly is pancreatic infiltrating ductal adenocarcinoma. These tumors arise in the head, body, or tail of the pancreas and are characterized by infiltrating desmoplastic stromal reactions (**Fig. 83-1B**).

Other subtypes of nonneuroendocrine pancreatic cancers include acinar cell carcinoma (tumors of the exocrine enzyme producing cell), medullary carcinoma, adenosquamous carcinoma, and other rare subtypes. Each of these is different in behavior and in their molecular characteristics and often requires other specific types of treatment.

Molecular Characteristics The molecular characteristics of pancreatic ductal adenocarcinoma reveal four genes that are commonly mutated or inactivated (sometimes referred to as the “four horsemen”). The most common is *KRAS* (usually in codon 12). It is critical to determine the specific mutation in *KRAS* because specific mutations may indicate specific therapies that should be considered. *KRAS* mutations are seen in virtually 100% of pancreatic adenocarcinomas. In fact, with the deep sequencing now available, if a *KRAS* mutation is not detected in the patient's tumor, one should consider that the tumor is likely of a different origin (e.g., small bowel, gallbladder, or cholangiocarcinoma—all of which could require different treatments). *p16/CDKN2A* is also noted in >90% of invasive pancreatic adenocarcinomas. *TP53* and *DPC4/MADH4* are mutated in about half of these tumors. As a reference point, the *BRCA2* gene noted in Table 83-1 is mutated in 7–10% of pancreatic adenocarcinomas.

Precursor Lesions Many pancreatic adenocarcinomas seem to arise from noninvasive epithelial precursor lesions. Detection of these could allow for early diagnosis of pancreatic cancer. These pancreatic intraepithelial neoplasias (PanINs) have varying degrees of dysplasia designated as PanINs 1–3 (and constitute a progression model for pancreatic cancer). Genetic alterations become more frequent as the PanIN grade increases (e.g., grade 3). Not all PanIN lesions progress to invasive malignancy. PanINs that are ≥ 1 cm are called *intraductal papillary neoplasms* and are usually noninvasive. If the intraductal tumor is in a branch duct, it is usually noninvasive; however, if the intraductal tumor is in a main duct and is large and nodular, it is more likely to have malignant behavior.

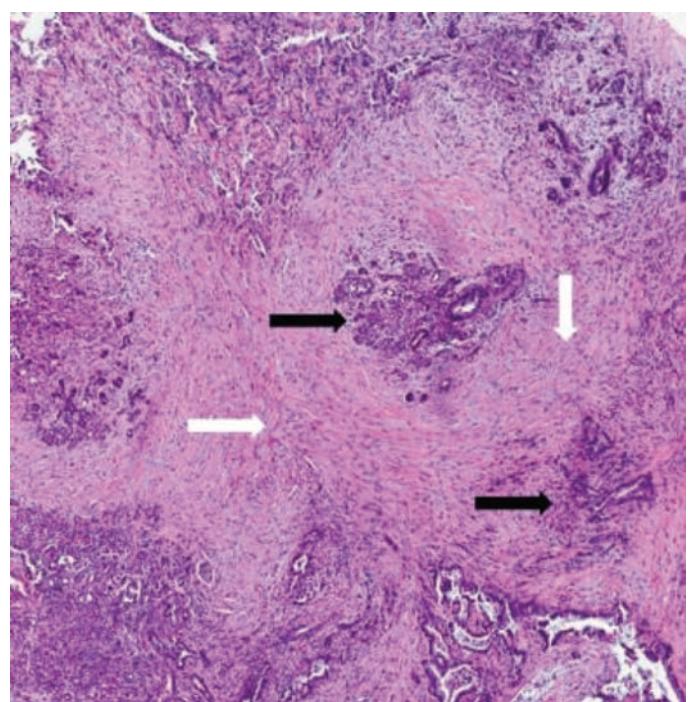
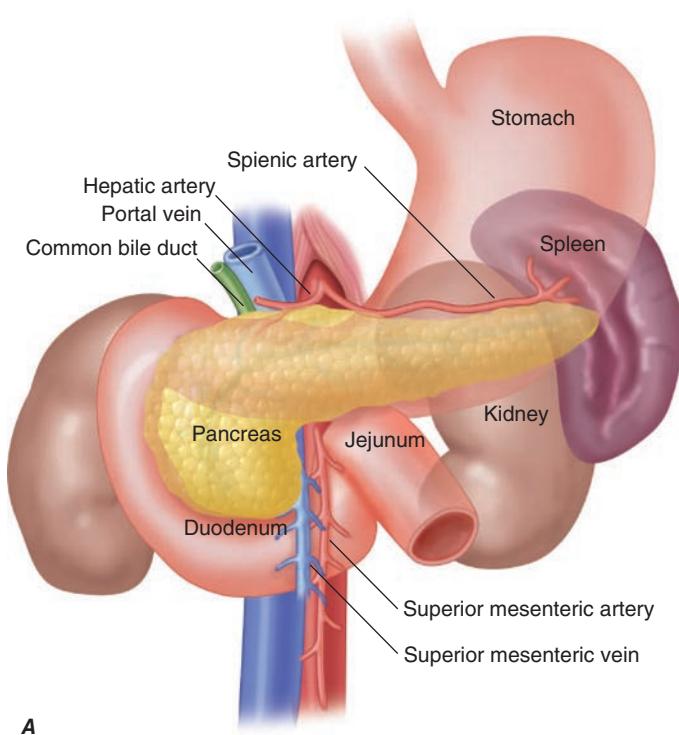


FIGURE 83-1 **A.** Note the relationship of the pancreas to the major vessels of the retroperitoneum. **B.** Ductal adenocarcinoma of the pancreas (black arrows), with intense stromal component (white arrows). (Part A is courtesy of Mary Kay Washington, MD, PhD, Vanderbilt University. Part B is courtesy of Haiyong Han, PhD, Translational Genomics Research Institute [TGen].)

One other pancreatic tumor is the mucinous cystic neoplasm; they may be seen as incidental findings on scans. These lesions are less likely invasive (20%) unless they are large and have nodules in them.

■ CLINICAL FEATURES

History and Physical The classic presentation for a patient with pancreatic cancer has been abdominal pain and weight loss with or without jaundice. The pain is midepigastric (sometimes described as a “boring-like” pain). Often the pain is in the back (due to retroperitoneal invasion of the splanchnic nerve plexus). The pain may be exacerbated by eating or lying flat. Other items of note in a history are light stool color from the absence of bile (steatorrhea also causes malodorous stools) and the onset of diabetes in the prior year. Jaundice, first detectable with a bilirubin of 2.5–3.0 mg/dL, is usually associated with tumor in the head of the pancreas. In some instances, depression is noted (with a higher subsequent number of suicides). Pruritis may be seen when the bilirubin reaches 6–8 mg/dL.

Physical signs include jaundice, signs of weight loss, a palpable gallbladder (Courvoisier’s sign), hepatomegaly, an abdominal mass, and even an enlarged spleen (usually indicating a portal vein thrombosis). Migratory superficial thrombophlebitis can also be seen (Trousseau’s syndrome). Signs of late disease include a lymph node palpable in the supraclavicular fossa (usually on the left where the thoracic duct enters the subclavian vein). This is clinically referred to as Virchow’s node. Occasionally, one can palpate subcutaneous metastases in the perumbilical area referred to as a Sister Mary Joseph’s node—named after one of the scrub nurses on the Mayo Clinic Operative Team who noted that when she prepped that area and felt those nodules, the patient often had peritoneal metastases.

The history and symptoms noted above may lead a person to see a physician; often CT and MRI scanning detects the disease before advanced disease symptoms appear.

■ DIAGNOSTIC WORKUP

Imaging Diagnostic imaging plays a major role in diagnosing pancreatic cancer and other intraabdominal diseases. The best technique

is the use of a dual-phase contrast-enhanced spiral CT using the pancreatic cancer protocol, which allows arterial phase enhancement and portal venous phase enhancement. This special protocol can provide helpful prospective staging and assessment of resectability. **Figure 83-2** demonstrates such a CT scan (with vascular involvement). **Figure 83-3** demonstrates the use of an 18F glucose positron emission tomography (PET) scan.

Histologic Diagnosis A histologic (tissue) diagnosis is essential and should be obtained with a cutting biopsy needle (not a skinny needle with cytology). Misdiagnosis is more common based on only fine-needle aspirates. Obtaining a tissue diagnosis allows not only for accuracy but also for molecular testing for KRAS mutations, microsatellite instability, and other important molecular abnormalities. Those molecular abnormalities and others will be increasingly important as more targeted therapies are developed for patients with pancreatic cancer.

The core needle (16–18 gauge) biopsy can be obtained via endoscopic ultrasound-guided techniques for a tumor localized to the pancreas or, if there are liver lesions or Virchow’s node, via percutaneous biopsy by interventional radiologists.

Serum Markers Before treatment, a serum sample should be obtained for levels of CA19-9, carcinoembryonic antigen (CEA), or if both are negative, for CA125 (can be positive when the CA19-9 is negative due to the patient not being a Lewis antigen secretor). These markers are not useful for staging but can be useful in following the course of pancreatic cancer.

■ IMPORTANT IMMEDIATE CONSIDERATIONS IN PATIENT CARE

While the patient is being evaluated and staged, one must be alert for biliary tract obstruction (and the attendant risk for sepsis from the biliary tree). A stent can be placed (plastic if temporary or metal if needed longer) to relieve the jaundice and pruritus. If surgery is being contemplated, an early surgical consultation is in order as some surgeons may want to proceed to surgery without placement of a stent. This immediate surgical approach is becoming less common as many multidisciplinary teams want consideration of use of chemotherapy

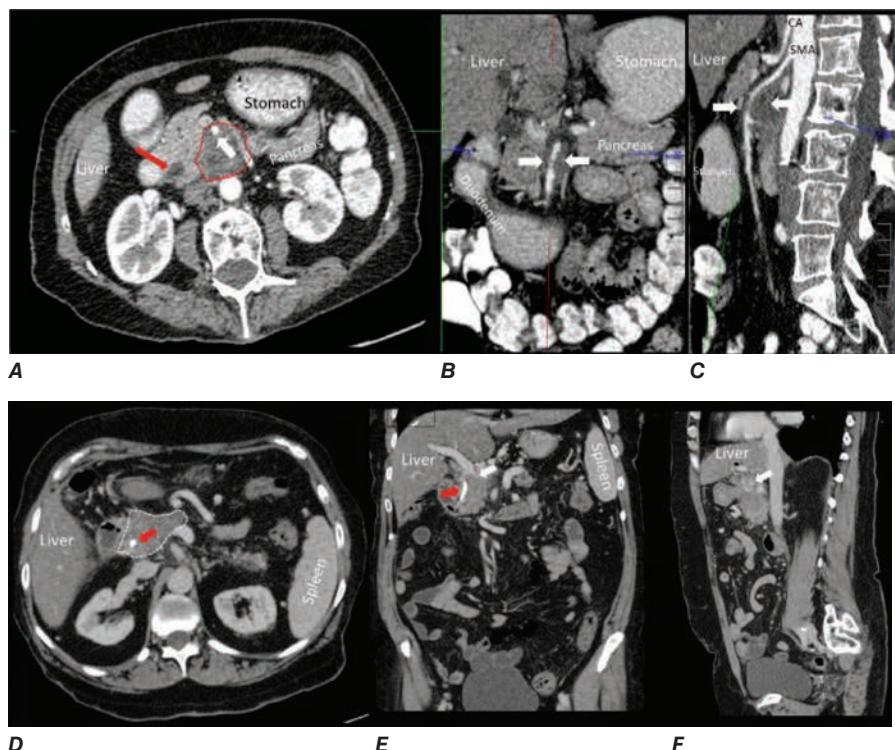


FIGURE 83-2 Selected images from contrast-enhanced CT in patients with locally advanced adenocarcinoma of the pancreas. A high-quality contrast-enhanced CT scan (arterial phase in panels **A–C** and portal venous phase in panels **D–F**) is required for optimal staging of pancreas cancer. Panel **A** demonstrates the typical features of adenocarcinoma of the pancreas on arterial phase axial CT scans (dotted outline) with tumor encasement of the superior mesenteric artery (white arrow). Note the dilatation of the common bile duct (red arrow). Panels **B** (magnified coronal) and **C** (sagittal) show reconstruction of CT images into additional orthogonal planes with exquisite details to confirm the unresectable nature of the tumor due to vascular encasement. Panel **D** demonstrates the typical features of adenocarcinoma of the pancreas on portal venous phase axial CT scans in a different subject. The dotted line outlines a pancreas cancer lesion in the pancreatic head, which is encasing the portal splenic confluence (dotted outline). Panels **E** (white arrow) and **F** show the pinched appearance of the portal splenic confluence by tumor abutment and invasion of the superior mesenteric vein (white arrow) on coronal and sagittal views. Note the presence of a stent in the common bile duct (red arrow) to help relieve biliary obstruction caused by the tumor. CA, celiac axis; SMA, superior mesenteric artery.

with or without radiation therapy (called neoadjuvant therapy) before a patient is taken to surgery.

Patients with pancreatic cancer are often hypercoagulable and frequently have migratory thrombophlebitis (Trousseau's sign) as well as deep vein thrombosis with pulmonary emboli (a frequent cause of

death). Appropriate examinations plus being alert to thromboses on the routine workup are mandatory so appropriate management can be put in place.

Control of pain or of any symptoms should be pursued to help patients be as comfortable as possible for their decision-making. Sometimes simple approaches like the use of a replacement pancreatic enzyme (at good therapeutic doses) can relieve the bloating, cramping, and diarrhea. Early involvement of a palliative care team can improve a patient's quality of life and sometimes even its length.

CLINICAL STAGING

The clinical staging of pancreatic cancer according to the American Joint Commission on cancer staging is presented in **Table 83-2**.

Table 83-3 presents another clinical way to express extent of disease as well as therapeutic approaches (to be discussed later).

For proper staging, some physicians believe that a laparoscopy either before or at the time of surgery is important. If metastatic disease is found at laparoscopy, one can avoid surgery that would not be helpful because disease is already advanced.

TREATMENT

Resectable Disease

Even for patients with resectable disease, the patient should be presented to a combined-modality conference. Some clinicians feel the best approach for patients with resectable disease (as defined in Table 83-3) is surgery. Only a small percentage of patients are in this category (10–20%). Some clinicians feel neoadjuvant therapy (chemotherapy before surgery) should be given before surgery (for controlling potential micrometastatic disease and shrinking the primary tumor). The surgery for patients with tumors in the head or

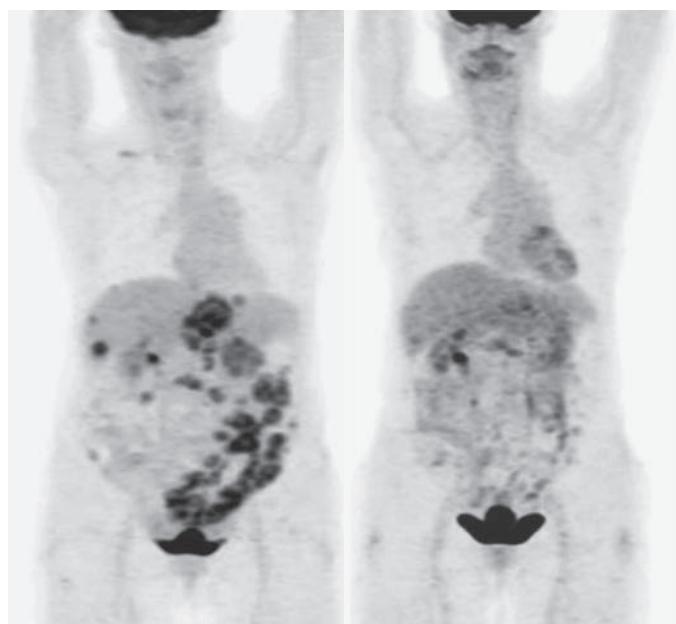


FIGURE 83-3 PET scan demonstrating metastatic disease—baseline and after 6 weeks of chemotherapy with some resolution of liver metastases.

TABLE 83-2 Definition of Primary Tumor (T)

T CATEGORY	T CRITERIA		
TX	Primary tumor cannot be assessed		
T0	No evidence of primary tumor		
Tis	Carcinoma in situ This includes high-grade pancreatic intraepithelial neoplasia (PanIn-3), intraductal papillary mucinous neoplasm with high-grade dysplasia, intraductal tubulopapillary neoplasm with high-grade dysplasia, and mucinous cystic neoplasm with high-grade dysplasia		
T1	Tumor ≤2 cm in greatest dimension		
T1a	Tumor ≤0.5 cm in greatest dimension		
T1b	Tumor >0.5 cm and <1 cm in greatest dimension		
T1c	Tumor 1–2 cm in greatest dimension		
T2	Tumor >2 cm and ≤4 cm in greatest dimension		
T3	Tumor >4 cm in greatest dimension		
T4	Tumor involves celiac axis, superior mesenteric artery, and/or common hepatic artery, regardless of size		
M CATEGORY	M CRITERIA		
M0	No distant metastasis		
M1	Distant metastasis		
N CATEGORY	N CRITERIA		
NX	Regional lymph nodes cannot be assessed		
N0	No regional lymph node metastases		
N1	Metastasis in one to three regional lymph nodes		
N2	Metastasis in four or more regional lymph nodes		
AJCC Prognostic Stage Groups			
WHEN T IS...	AND N IS...	AND M IS...	THEN THE STAGE GROUP IS....
Tis	N0	M0	0
T1	N0	M0	IA
T1	N1	M0	IIB
T1	N2	M0	III
T2	N0	M0	IB
T2	N1	M0	IIB
T2	N2	M0	III
T3	N0	M0	IIA
T3	N1	M0	IIB
T3	N2	M0	III
T4	Any N	M0	III
Any T	Any N	M1	IV

Source: Used with the permission of American College of Surgeons. MB Amin et al (eds): *AJCC Cancer Staging Manual*, 8th ed. Springer, 2017.

uncinate body of the pancreas is usually a pylorus-sparing pancreaticoduodenectomy (a modified Whipple procedure). For tumors in the body or tail, a distal pancreatectomy is usually performed. Clinical and pathologic findings of the resection are defined as either an R0 resection (no macroscopic or microscopic disease left

after surgery) or an R1 resection, which refers to residual disease likely left behind. Patients with smaller tumors and lymph node-negative disease have a better survival (median of about 18–23 months with 5-year survival of about 20%).

Two approaches are being explored to try to improve on this.

- Postoperative adjuvant therapy. The standard of care is to use 24 weeks of adjuvant treatment with a modified folinic acid, 5-fluorouracil, irinotecan, and oxaliplatin (FOLFIRINOX) regimen. In the definitive clinical trial, the median survival was 54 months for the combination of modified FOLFIRINOX versus 35 months for gemcitabine alone (hazard ratio [HR] 0.64; 95% confidence interval [CI] 0.48–0.86; $p = .003$). Toxicities were manageable.
- A newer approach is the use of neoadjuvant chemotherapy (chemotherapy given before surgery) to try to shrink the tumor and normalize the patient's serum CA19-9 level. Data suggest that patients who have borderline resectable/locally advanced disease can benefit from neoadjuvant therapy. Studies of neoadjuvant chemotherapy with or without radiation therapy are ongoing.

LOCALLY ADVANCED DISEASE (30% OF PATIENTS)

For patients with locally advanced disease, the median survival is also quite poor (6–10 months) because many of the patients die with local problems (e.g., portal vein thrombosis with bleeding

TABLE 83-3 Extent of Disease and Therapeutic Approach

DESIGNATION (MEDIAN SURVIVAL)	THERAPEUTIC APPROACHES
1. Resectable (localized): (18–23 mo)	Surgical option (or preoperative-neoadjuvant therapy first) Surgery is followed by postsurgery adjuvant therapy • Currently mFOLFIRINOX
2. Locally advanced: (6–10 mo)	Either chemotherapy or chemotherapy + radiation therapy • Encasement of arteries • Venous occlusion (superior mesenteric vein [SMV] or portal) • No extrapancreatic disease
3. Metastatic: (8.3–12.8 mo)	Systemic chemotherapy

Abbreviation: mFOLFIRINOX, modified FOLFIRINOX (folinic acid, 5-fluorouracil, irinotecan, and oxaliplatin (T Conroy et al: *N Engl J Med* 379:2395, 2018).

varices, obstruction, sepsis). The approach has been to try to reduce the bulk of the disease with use of radiation therapy plus chemotherapy or chemotherapy alone, with the goal that the disease could become resectable. No standard therapy has been agreed upon, but experimental approaches are applying some of the treatments that show promise in advanced metastatic disease.

ADVANCED METASTATIC DISEASE (60% OF PATIENTS)

Only a few of the many phase 3 randomized trials in patients with advanced pancreatic cancer have led to meaningful increases in survival. We have learned that a regimen needs to have at least a 50% improvement in overall survival or 90% improvement in 1-year survival in a pilot phase 2 trial to predict for any degree of success in large randomized phase 3 trials.

Patients with the best chance of receiving a benefit from treatment have a good performance status (functioning up and around at least 70% of the day), have a reasonable albumin level (≥ 3.0 g/dL), and a neutrophil/lymphocyte ratio of ≤ 5.0 .

Single-agent gemcitabine achieves a median survival of 6 months and a 1-year survival rate of 18%. **Table 83-4** details three combination regimens that have further improved survival modestly. Median overall survival still ranges from 6 to 11 months. However, 1-year survival is now approaching 35% for these combination regimens with some long-term 4+ year survivors.

Also of note in Table 83-4, liposomal irinotecan has U.S. Food and Drug Administration (FDA) approval in combination with 5-fluorouracil and leucovorin for patients whose tumors have progressed on gemcitabine (e.g., second-line therapy for stage IV disease) based on improved overall survival.

FOR PATIENTS WITH A SPECIFIC MOLECULAR PROFILE IN THEIR TUMOR/GERMLINE

PARP inhibitors have clinical activity against pancreatic cancers having mutations in *BRCA2*, *BRCA1*, or *PALB2* (i.e., defective DNA repair proteins). In addition, their tumors might be more sensitive to specific combinations of chemotherapy like gemcitabine plus cisplatin. In addition, tumors with microsatellite instability often have more mutations, and such tumors appear to have a higher response

TABLE 83-4 Combination Chemotherapy Regimens That Have an Impact on Survival

STUDY DESIGN (AUTHOR/REF)	NO. OF PATIENTS	MEDIAN SURVIVAL (MONTHS)
Gemcitabine + erlotinib vs gemcitabine (Moore et al: J Clin Oncol 26:1960, 2007)	569	6.24 vs 5.91 (HR 0.82; 95% CI 0.69–0.99; $p = .038$)
FOLFIRINOX (folinic acid + 5-fluorouracil + irinotecan + oxaliplatin) vs gemcitabine (Conroy et al: N Engl J Med 364:1817, 2011)	342	11.1 vs 6.8 (HR 0.57; 95% CI 0.45–0.70; $p < .001$)
Nap-paclitaxel + gemcitabine vs gemcitabine, (Von Hoff et al: N Eng J Med 369:1691, 2013.)	861	8.5 vs 6.7 (HR 0.72; 95% CI 0.62–0.83; $p < .001^a$)
Nanoliposomal irinotecan + fluorouracil + folinic acid vs nanoliposomal irinotecan monotherapy vs fluorouracil + folinic acid (Wang-Gillam et al: Lancet 387:545, 2015)	417	6.1 vs 4.2 (HR 0.67; 95% CI 0.49–0.92; $p = .012^b$)

^aThe 2-year survival rate with this regimen is 9% and the 3+ year rate is 4%. Other studies have not reported on these parameters. ^bHR is for nanoliposomal irinotecan + 5-fluorouracil + folinic acid vs 5-fluorouracil + folinic acid.

Abbreviations: CI, confidence interval; HR, hazard ratio.

rate to immunotherapy with checkpoint inhibitors and anti-PD-1 (pembrolizumab, nivolumab) and anti-PD-L1 antibodies.

MAINTENANCE THERAPY FOR PATIENTS RESPONDING TO TREATMENT

For patients with germline *BRCA1* or *BRCA2* mutations whose disease has not progressed during a first-line platinum-based regimen, the PARP inhibitor olaparib has been shown to improve progression-free survival (7.4 vs 3.8 months; HR 0.53; 95% CI 0.35–0.82; $p = .004$) with no change in quality of life.

OTHER POTENTIAL FACTORS INFLUENCING SURVIVAL

Preclinical studies have suggested that vitamin D can inhibit the development and growth of cancer. In models of pancreatic cancer, synthetic analogues of vitamin D had an effect on both tumor cells and on the tumor microenvironment. Clinical studies are conflicting as to whether circulating levels of plasma 25-hydroxyvitamin D (25[OH]D) affect the incidence of pancreatic cancer. However, patients with prediagnostic levels of 25(OH)D that are in the normal range have a longer survival than those who have reduced levels (35% lower hazard for death).

■ FUTURE DIRECTIONS

Death from pancreatic cancer is often due to progressive inanition. The metabolic consequences of this cancer are being examined. The tumor can be fatal at a modest level of tumor burden based on the profound metabolic effects. Other promising areas of investigation include addressing the florid stromal reaction around the tumor cells (believed to act as a physical barrier to drug delivery and as an immune sanctuary for the tumor cells). Improvements in outcomes for pancreatic cancer would accompany earlier detection. A small decrease in the percentage of patients being diagnosed with stage IV pancreatic cancer has been noted. The reason for this encouraging sign is unknown. The 5-year survival for earlier stage patients has increased from 44.7% in 2004 to 83.7% in 2012. There is emerging evidence that the surveillance to detect *CDKN2A* mutation carriers can detect pancreatic ductal cancer at a resectable stage.

ACKNOWLEDGMENT

Thank you to Nicole Harkey, for assistance in the preparation of this chapter, and Drs. Elizabeth Washington, Ron Korn, and Haiyong Han and the American Joint Committee on Cancer for providing the figures and tables.

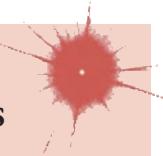
■ FURTHER READING

- CONROY T et al: FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med* 364:1817, 2011.
- CONROY T et al: FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. *N Engl J Med* 379:2395, 2018.
- GOLAN T et al: Maintenance olaparib for germline BRCA-mutated metastatic pancreatic cancer. *N Engl J Med* 381:317, 2019.
- HRUBAN RJ et al: Genetic progression in the pancreatic ducts. *Am J Pathol* 156:1821, 2000.
- RAHIB L et al: Evaluation of pancreatic cancer clinical trials and benchmarks for clinically meaningful future trials: A systemic review. *JAMA Oncol* 2:1209, 2016.
- RAWLA P et al: Epidemiology of pancreatic cancer: Global trends, etiology and risk factors. *World J Oncol* 10:10, 2019.
- SOLOMON S et al: Inherited pancreatic cancer syndromes. *Cancer J* 18:485, 2012.
- VON HOFF D et al: Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *N Engl J Med* 369:1691, 2013.

84

Gastrointestinal Neuroendocrine Tumors

Matthew H. Kulke



Gastrointestinal (GI) neuroendocrine tumors (NETs) can be broadly grouped according to their site of origin, as either extrapancreatic NETs, historically called carcinoid tumors, or pancreatic NETs. While NETs can pursue a broad range of clinical behaviors, they classically follow a course that is more indolent than many other malignancies. NETs also have the ability to synthesize peptides, growth factors, and bioactive amines that may be ectopically secreted, giving rise to a range of unique clinical syndromes.

INCIDENCE AND PREVALENCE

The diagnosed incidence of NETs has been steadily increasing over the past several decades (Fig. 84-1). An analysis of data from the Surveillance, Epidemiology, and End Results (SEER) program, comprising population-based data in the United States from 1973 to 2012, showed that the incidence had increased 6.4-fold over this time period and that the estimated prevalence of patients who had been diagnosed with a NET was >170,000. This analysis also found that overall survival durations for patients with NETs had improved significantly. The increasing incidence and improved survival durations for patients with NETs likely reflect, at least in part, advances in both diagnosis and treatment. While environmental or other factors leading to an increased incidence of NETs cannot be excluded; common cancer risk factors such as tobacco or alcohol use and dietary patterns have not been clearly linked to NET development.

A minority of NETs develop in the context of autosomal inherited genetic syndromes associated with mutations in specific tumor-suppressor genes. The most common of these is multiple endocrine neoplasia type 1 (MEN 1), due to mutation and loss of function of the *menin* gene, located on chromosome 11q13 (Chap. 388). Patients with MEN 1 are at risk for developing pancreatic NETs as well as hyperparathyroidism and pituitary adenomas; less commonly, they may develop bronchial and thymic NETs. Other inherited syndromes

associated with NETs include von Hippel-Lindau disease (VHL), von Recklinghausen's disease (neurofibromatosis type 1), and tuberous sclerosis (Bourneville's disease). Inherited mutations in the VHL gene, located on chromosome 3p25, are associated with the development of cerebellar hemangioblastomas, renal cancer, and pheochromocytomas and, less commonly, pancreatic NETs. Mutations in neurofibromin (*NF1*) are associated with neurofibromatosis (von Recklinghausen's disease); patients with neurofibromatosis are at risk of developing both pancreatic and extrapancreatic NETs. Tuberous sclerosis is caused by mutations that alter either hamartin (*TSC1*) or tuberin (*TSC2*). Both hamartin and tuberin function as inhibitors of the phosphatidylinositol 3-kinase and the mechanistic target of rapamycin (mTOR) signaling cascades, and pancreatic NETs have been reported in these patients. Rare cases of familial small intestine NETs have also been reported; in these cases, multiple synchronous tumors generally arise within the small intestine. A characteristic inherited mutation, however, has not been identified to date in the majority of these cases.

HISTOLOGIC CLASSIFICATION AND MOLECULAR FEATURES

The histologic features of NETs vary widely and are one of the most important determinants of both clinical behavior and treatment. NETs are classified based on the degree tumor differentiation (well or poorly differentiated), as assessed by a pathologist, and tumor grade (grades 1–3) (Table 84-1). Tumor grade closely correlates with mitotic count and Ki-67 proliferative index. Classic, well-differentiated NETs are composed of monotonous sheets of small round cells with uniform nuclei and only rare mitoses. Immunocytochemical staining for chromogranins and synaptophysin is typical. Ultrastructurally, these tumors contain electron-dense neurosecretory granules containing peptides and bioactive amines that may be ectopically secreted, giving rise to a range of clinical syndromes. These classic well-differentiated NETs have low-grade features and generally have a mitotic index of <2 mitoses per 10 high-power field (HPF) and a Ki-67 proliferative index of <3%. Less commonly, well-differentiated NETs have an intermediate histologic grade and pursue a somewhat more aggressive clinical course. These intermediate-grade tumors typically have a mitotic count of 2–20 per 10 HPF and a mitotic index of 3–20%. Well-differentiated high-grade tumors are rare and have mitotic counts that exceed 20 per 10 HPF and a Ki-67 proliferative index of >20%. Poorly

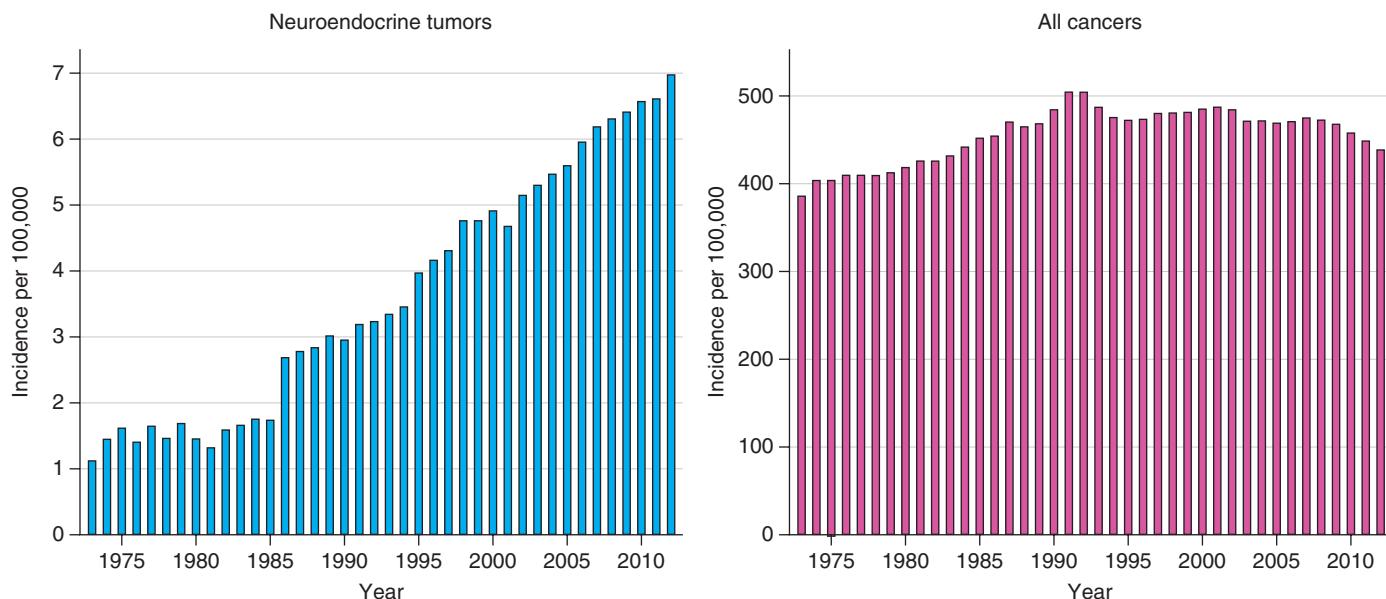


FIGURE 84-1 Incidence of neuroendocrine tumors (NETs). The incidence of neuroendocrine tumors has been increasing over the past several decades, an observation that has been attributed in part to improved diagnosis and classification. (Adapted from A Dasari et al: Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol* 3:1335, 2017.)

TABLE 84-1 Histologic Classification of Neuroendocrine Tumors

CLASSIFICATION	DIFFERENTIATION	GRADE	MITOTIC COUNT	KI-67
Neuroendocrine tumor	Well differentiated	Low grade (grade 1)	<2 per 10 HPF	<3%
Neuroendocrine tumor	Well differentiated	Intermediate grade (grade 2)	2–20 per 10 HPF	3–20%
Neuroendocrine tumor	Well differentiated	High grade (grade 3)	>20 per 10 HPF	>20%
Neuroendocrine carcinoma	Poorly differentiated	High grade (grade 3)	>20 per 10 HPF	>20%

Abbreviation: HPF, high-power field.

differentiated high-grade tumors form the most clinically aggressive category; prognosis and treatment for these tumors differ markedly from their well-differentiated counterparts.

Whole exome sequencing of sporadic pancreatic NETs found that the most frequently altered gene was *MEN1*, occurring in 44% of tumors. In addition, 43% of tumors had mutations in genes encoding two subunits of a transcription/chromatin remodeling complex consisting of DAXX (death-domain-associated protein) and ATRX (α-thalassemia/mental retardation syndrome X-linked). Mutations in genes associated with the mTOR pathway were identified in 15% of tumors. In contrast, recurrent mutations in extrapancreatic NETs appear to be rare. In one study that evaluated 180 small intestinal NETs using a combination of whole exome and more targeted genome-sequencing analysis, recurrent mutations were only observed in the *CDKN1B* gene (cyclin-dependent kinase inhibitor 1B [p27^{Kip1}]) in 8% of cases. Loss of chromosome 18 is a common finding in small-bowel NETs. Small-intestinal GI carcinoids commonly have epigenetic changes; however, the clinical significance of these alterations remains uncertain. Initial studies have suggested that well-differentiated pancreatic and extrapancreatic NETs express only low levels of the immune checkpoint markers PD-1 and PD-L1.

CLINICAL PRESENTATION AND MANAGEMENT OF LOCALIZED PANCREATIC NEUROENDOCRINE TUMORS

Pancreatic NETs have been subcategorized as either “functional,” meaning associated with symptoms of hormone secretion, or non-functional, in which case they may be clinically silent until they cause anatomic symptoms. The clinical presentation of functional pancreatic NETs depends on the type of hormone secreted and can sometimes lead to dramatic clinical presentations (Table 84-2). The most common functional pancreatic NETs are insulinomas, followed in incidence by glucagonomas and gastrinomas. Pancreatic NETs secreting other hormones, including somatostatin, vasoactive intestinal peptide (VIP), adrenocorticotrophic hormone (ACTH), and parathyroid hormone (PTH) have also been described but are uncommon. Only ~20% of pancreatic NETs are associated with symptoms of hormone hypersecretion; the majority of pancreatic NETs are “nonfunctional” and are diagnosed either incidentally or after patients present with abdominal pain, weight loss, or other anatomic symptoms related to tumor bulk.

GASTRINOMA

Patients with gastrinoma typically present with Zollinger-Ellison syndrome (ZES) (Chap. 324). The most common symptoms associated with this syndrome are abdominal pain, diarrhea, gastroesophageal reflux disease (GERD), and peptic ulcer disease. Peptic ulcer disease manifesting as multiple ulcers with associated diarrhea is a classic presentation. Up to 25% of patients with ZES have MEN 1, and a diagnosis of gastrinoma should prompt a family history as well as an assessment for concurrent hyperparathyroidism. Fasting hypergastrinemia is a nearly universal finding in patients with gastrinoma. Importantly, however, proton pump inhibitors (PPIs) can suppress acid secretion sufficiently to cause hypergastrinemia and can confound the diagnosis. Achlorhydria, usually in the context of chronic atrophic gastritis, will also elevate serum gastrin levels but can usually be easily distinguished from gastrinoma given the absence of other evidence of acid hypersecretion.

While often classified as pancreatic NETs, the majority of gastrinomas in fact arise in the “gastrinoma triangle,” an anatomic region

bounded by the duodenum, pancreas, and confluence of the cystic and common bile ducts. Most gastrinomas (50–90%) in sporadic ZES arise in the duodenum. They are frequently small and may be difficult to localize. Imaging studies generally include either CT or MRI; endoscopic ultrasound or somatostatin scintigraphy may also be helpful.

PPIs are generally highly effective in the treatment of symptoms related to gastrinoma and are considered the initial treatment of choice. Rapid resolution of both abdominal pain and diarrhea related to acid hypersecretion is common. Somatostatin analogues may also be helpful in controlling symptoms in refractory cases. Once symptoms are controlled, surgical resection is generally recommended for patients with sporadic gastrinomas, both to eliminate the cause of gastrin secretion and to decrease the risk of developing metastatic disease. The technique used for resection depends in large part on the precise location of the tumor. In some cases where preoperative imaging is not successful but a diagnosis is strongly suspected, exploratory laparotomy with intraoperative ultrasound may be undertaken. In gastrinoma patients who have underlying MEN 1, tumors are generally small and multiple; the role of routine surgery in this setting remains more controversial but generally is still recommended in patients with larger tumors measuring ≥1.5–2 cm in diameter.

INSULINOMA

Patients with insulinoma generally present with symptoms of hypoglycemia, which may include confusion, headache, disorientation, visual difficulties, irrational behavior, and even coma. In some cases, the diagnosis of insulinoma may not be immediately evident, and patients with

TABLE 84-2 Clinical Presentation and Management of Secretory Syndromes Associated with Neuroendocrine Tumors

	CLINICAL SYMPTOMS AND MANIFESTATIONS	TREATMENT OPTIONS TO CONTROL SECRETORY SYMPTOMS
Pancreatic Neuroendocrine Tumors		
Gastrinoma (generally located in “gastrinoma triangle”)	Zollinger-Ellison syndrome: gastroesophageal reflux, peptic ulcer disease, diarrhea	Proton pump inhibitors, somatostatin analogues
Insulinoma	Hypoglycemia leading to confusion, lethargy, coma; weight gain	Diazoxide, everolimus
Glucagonoma	Skin rash (necrolytic migratory erythema), glucose intolerance, weight loss	Somatostatin analogues
VIPoma	Verner-Morrison syndrome: watery diarrhea, hypokalemia, achlorhydria	Somatostatin analogues
ACTHoma	Cushing’s syndrome: hyperglycemia, weight gain, hypokalemia	Ketoconazole, consider adrenalectomy
Extrapancreatic gastrointestinal neuroendocrine tumors		
Typically in setting of advanced disease from small intestine or appendiceal primary tumors	Carcinoid syndrome: flushing, diarrhea, right-sided valvular heart disease, mesenteric fibrosis	Somatostatin analogues, telotristat ethyl

insulinoma may initially be diagnosed with psychiatric illnesses that in retrospect were hypoglycemic symptoms. The diagnosis of insulinoma is generally confirmed with elevated fasting insulin levels in conjunction with elevated proinsulin and C-peptide. Fasting hypoglycemia can also be caused by severe liver disease, alcoholism, and poor nutrition. Postprandial hypoglycemia may also occur after gastric bypass surgery. Surreptitious use of insulin or hypoglycemic agents may be difficult to distinguish from an insulinoma. Evaluation of proinsulin and C-peptide levels, both of which should be normal in patients using exogenous insulin, and measurement of sulfonylurea levels in serum or plasma are helpful in such cases.

The hypoglycemia associated with insulinomas can be severe and challenging to manage. Diazoxide has historically been used in the initial management of patients with insulinoma and results in inhibition of insulin release, though it can also be associated with side effects including sodium retention and nausea. Everolimus, in addition to its antitumor effect (see below), is highly effective in improving glycemic control in patients with insulinoma. The benefits of everolimus in this setting may be related both to induction of insulin resistance and a direct antitumor effect. While somatostatin analogues are usually effective in treating symptoms of hormone hypersecretion associated with other types of NETs, they should be used with caution in patients with insulinoma. Somatostatin analogues may suppress counterregulatory hormones, such as growth hormone (GH), glucagon, and catecholamines, and precipitously worsen hypoglycemia.

Insulinomas may be difficult to localize, as they are less consistently avid on somatostatin scintigraphy than other pancreatic NETs. Insulinomas are also generally small, with the majority measuring <2 cm in diameter. Because of their generally small size, insulinomas are best localized with endoscopic ultrasound (EUS). In the absence of metastatic disease, surgical resection is usually recommended. The primary treatment for exophytic or peripheral insulinomas is enucleation. If enucleation is not possible because of invasion or the location of the tumor within the pancreas, then pancreateoduodenectomy for tumors in the head of the pancreas or distal pancreatectomy with preservation of the spleen for smaller tumors not involving splenic vessels may be considered.

■ GLUCAGONOMA

Patients with glucagonoma most commonly present with a characteristic dermatitis, called necrolytic migratory erythema (Fig. 84-2). The rash usually involves intertriginous sites, especially in the groin or buttock, and can wax and wane. Other common presenting symptoms of glucagonoma include glucose intolerance and weight loss. The diagnosis of glucagonoma can be confirmed by demonstrating an increased plasma glucagon level, generally in excess of 1000 pg/mL. Somatostatin analogues are usually highly effective as an initial treatment to alleviate the symptoms and rash associated with glucagon hypersecretion. The majority of glucagonomas are large in size at presentation and arise in the tail of the pancreas. For patients with localized disease, distal pancreatectomy and splenectomy are recommended. A hypercoagulable state has been reported in up to 33% of patients with glucagonoma, and perioperative anticoagulation should generally be employed.

■ SOMATOSTATINOMA

Patients with somatostatinoma typically present with diabetes mellitus, gallbladder disease, diarrhea, and steatorrhea. Somatostatinomas occur primarily in the pancreas or duodenum, are usually large, and are commonly metastatic at presentation. They are only rarely associated with MEN 1. The diagnosis of somatostatinoma is based on the demonstration of elevated plasma somatostatin levels, and as such, the potential benefits of using somatostatin analogs as a treatment for patients with somatostatinoma is questionable. Surgery is recommended for patients with localized disease.

■ VIPOMA

VIPomas are associated with a distinct syndrome that has been variously called Verner-Morrison syndrome, pancreatic cholera, and WDHA syndrome (watery diarrhea, hypokalemia, and achlorhydria).



FIGURE 84-2 Glucagonoma syndrome. Patients with glucagonoma may present with a classic skin rash, necrolytic migratory erythema (shown). Other presenting symptoms include glucose intolerance and weight loss.

VIP is a 28-amino-acid peptide that mimics the effects of the cholera toxin by stimulating chloride secretion in the small intestine and increasing smooth-muscle contractility, resulting in profound diarrhea. Treatment of dehydration, hypokalemia, and electrolyte losses with fluid and electrolyte replacement is the most critical initial treatment for patients with VIPoma. VIPomas are usually solitary and arise in the pancreatic tail. Elevated plasma levels of VIP are typical but should not be the only basis of the diagnosis of VIPomas because they can occur with some diarrheal states including inflammatory bowel disease, in the setting of small bowel resection, and radiation enteritis. Chronic surreptitious use of laxatives/diuretics can be particularly difficult to detect clinically. Somatostatin analogues are usually highly effective in controlling the diarrhea; surgical resection is recommended for patients with localized disease.

■ OTHER SECRETORY PANCREATIC NETS

Pancreatic NETs secreting GH-releasing factor (GRF), calcitonin, ACTH, and PTH-related protein have also been described; it is also possible for pancreatic NETs to secrete more than one hormone or for the secretory profiles to evolve over time. Gastrinomas, in particular, may evolve and may be associated with secretion of ACTH, resulting in ectopic Cushing's syndrome. Tumors secreting these hormones may not be as responsive to treatment with somatostatin analogues as the more common pancreatic NETs and the associated hormonal symptoms may cause significant morbidity. As with other pancreatic NETs, patients with localized disease are generally treated with surgical resection. In patients with ACTH-secreting tumors, the associated symptoms of Cushing's syndrome can be alleviated with adrenalectomy if resection of the primary tumor is not possible or in the setting of metastatic disease.

■ PANCREATIC NETS ARISING IN THE SETTING OF MEN 1

Pancreatic NETs occurring in patients with MEN 1 are typically multiple and often pursue a relatively indolent course. Because of the high probability of multiple tumors, surgical resection of confirmed pancreatic NETs in patients with MEN 1 is usually undertaken with caution given the likelihood of tumors arising in the remaining pancreas if

partial pancreatectomy is undertaken as well as the significant morbidities associated with total pancreatectomy. However, for symptomatic tumors or for growing tumors >2 cm in size, surgical resection may still be considered.

■ NONFUNCTIONING PANCREATIC NETS

As noted above, the majority of pancreatic NETs are not associated with symptoms of hormone hypersecretion and are considered “non-functional.” As a result, they often remain clinically silent and either are diagnosed incidentally or are not diagnosed until widespread, metastatic disease is present resulting in anatomic symptoms. If they are localized at diagnosis, the general treatment recommendation is surgical resection; however, the management of small, asymptomatic pancreatic NETs is debated. Assuming tumors are low grade, patients with incidentally discovered, low-grade tumors measuring <1 cm in size can be safely followed; other retrospective studies have suggested nonoperative management for nonfunctioning pancreatic NETs measuring up to 3 cm. In contrast, however, an analysis of the SEER database suggested that at least some tumors measuring <2 cm in size can pursue a more aggressive course. Management of small, incidentally discovered, asymptomatic, low-grade pancreatic NETs is therefore based on clinical judgement, taking into account surgical risk and patient comorbidities.

CLINICAL PRESENTATION AND MANAGEMENT OF LOCALIZED EXTRAPANCREATIC GASTROINTESTINAL NEUROENDOCRINE TUMORS

Extrapancreatic GI NETs, historically called carcinoid tumors, may arise virtually anywhere in the GI tract and differ significantly in their clinical characteristics depending on their location. The most common locations for extrapancreatic NETs are the stomach, distal small intestine, appendix, and rectum.

■ GASTRIC NETS

Gastric NETs can be categorized into three groups: type 1 (associated with chronic atrophic gastritis); type 2 (associated with gastrinomas and ZES), and type 3 (sporadic, gastric NETs). Type 1 gastric NETs are the most common of the three types. In type 1 gastric NETs, chronic atrophic gastritis results in loss of acid secretion with consequent loss of the negative feedback loop on gastrin-producing cells in the antrum of the stomach. Pernicious anemia is also commonly associated with this condition; classic laboratory findings are a markedly elevated gastrin level and low levels of vitamin B₁₂. Unchecked gastrin secretion in these patients results in hyperplasia of the endocrine cells in the gastric fundus. A typical finding on endoscopy is diffuse endocrine cell hyperplasia with multiple gastric carcinoid tumors (Fig. 84-3). These tumors generally pursue a benign course and can be monitored with serial endoscopy. In cases where tumors continue to grow or become symptomatic, antrectomy to remove the source of gastrin production

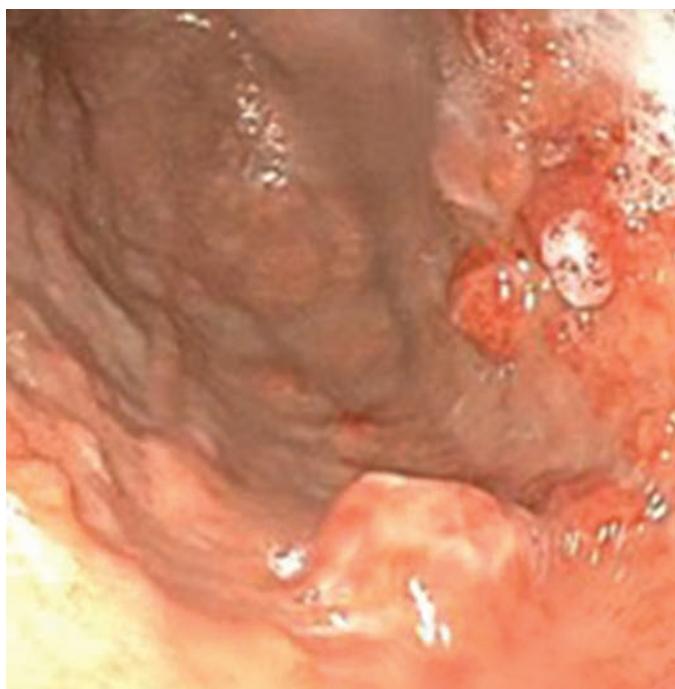


FIGURE 84-3 Multifocal gastric neuroendocrine tumor. (Courtesy of Christopher Huang MD, Boston Medical Center.)

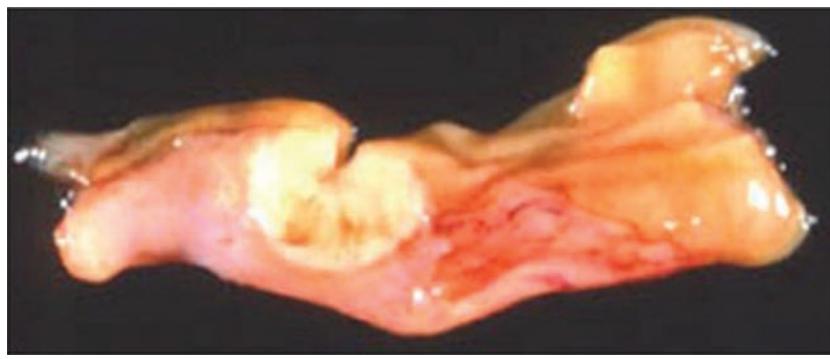
can result in tumor regression. Type 2 tumors are rare and usually occur in the setting of gastrinoma; as with type 1 gastric NETs, elevated gastrin levels result in diffuse gastric neuroendocrine hyperplasia and multifocal gastric NETs. Resection of the gastrinoma, removing the source of gastrin production, is the treatment of choice.

In contrast to type 1 and type 2 gastric NETs, type 3 gastric NETs are generally solitary, arise in the setting of normal gastrin levels, and may pursue a far more aggressive course. For early-stage, smaller tumors, endoscopic or wedge resection may be performed. For larger tumors, partial gastrectomy with lymphadenectomy is recommended.

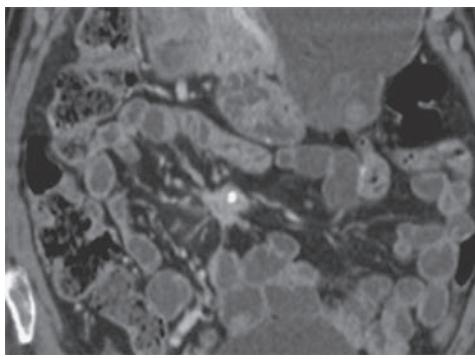
■ NETS OF THE SMALL INTESTINE

Small-bowel NETs occur most commonly in the terminal ileum and are notoriously difficult to diagnose at an early stage. One reason for this is that they arise within the muscularis, and their submucosal location makes them difficult to see during routine colonoscopy (Fig. 84-4A). Small-bowel NETs are also often multifocal; multifocal tumors appear to arise independently throughout the small intestine, although the mechanisms underlying this phenomenon remain uncertain.

Small-bowel NETs are often associated with desmoplasia and mesenteric fibrosis, likely as a result of fibroblast proliferation stimulated by tumor serotonin secretion. Mesenteric fibrosis frequently



A



B

FIGURE 84-4 Small intestine neuroendocrine tumor. **A.** Small intestine neuroendocrine tumors arising in submucosal location. The submucosal location of small intestine neuroendocrine tumors, together with their location beyond the ileocecal valve in the terminal ileum, can make endoscopic detection challenging. **B.** Classic “spoke and wheel” appearance of calcified mesenteric mass associated with small intestine primary neuroendocrine tumor. Mesenteric fibrosis commonly leads to intermittent obstructive symptoms and can also lead to ischemia when the mesenteric vasculature is involved. (Fig. B: Courtesy of Christina LeBedis MD, Boston Medical Center.)

results in intermittent small-bowel obstruction and, in some cases, bowel ischemia due to involvement of the mesenteric vessels. Patients may experience symptoms of intermittent abdominal pain and associated diarrhea, sometimes for months or years before diagnosis, that because of the difficulty in diagnosis are often attributed to irritable bowel syndrome. One classic finding that can aid in diagnosis is that the lymph node metastases associated with small intestine NETs are usually larger than the primary tumor and may be calcified, which, together with the tethering of the small intestine caused by the associated fibrosis, results in a classic “spoke and wheel” appearance on computed tomography (Fig. 84-4B).

Surgical resection of the primary tumor and associated metastases is recommended when feasible and is performed with curative intent when distant metastatic disease is not already present. Resection should also be considered in patients with metastatic disease experiencing intermittent obstruction or abdominal discomfort thought to be related to the primary tumor or associated mesenteric disease. Some have also advocated the routine resection of asymptomatic small-bowel primary tumors in patients with distant metastatic disease, with the rationale that this may be a way to prevent the future development of fibrosis and obstruction and preempt the development of unresectable disease due to tumor involvement of the mesenteric vessels. However, the available data on the benefits of resecting an asymptomatic primary tumor in this context are conflicting. Some studies have suggested that this practice results in an overall survival benefit, but the retrospective nature of these studies makes the data difficult to interpret given the high potential for selection bias in patients taken to surgery compared with those who were not. Other studies have suggested that prophylactic primary tumor resection confers no survival benefit and that surgery can be safely delayed until it is indicated based on the development of symptoms.

■ NETS OF THE APPENDIX

NETs are one of the most common tumors arising in the appendix. They are typically discovered incidentally in younger individuals undergoing appendectomy for acute appendicitis and not uncommonly are identified only at the time of pathology review. While the unexpected diagnosis of an appendiceal NET in such situations can cause considerable anxiety, in the majority of cases, the prognosis is excellent. Indeed, the clinical behavior of appendiceal NETs has been inferred from multiple large retrospective surgical series that suggest that the risk of lymph node or distant metastases from appendiceal NETs with well-differentiated histology and a tumor diameter measuring <2 cm is extremely low. In such cases, appendectomy alone is felt to be a sufficient surgical procedure.

In contrast, the risk of metastases for tumors measuring 2–3 cm is ~20–30% and is even greater for tumors measuring >3 cm. For patients with larger tumors, more formal staging studies with either cross-sectional imaging or somatostatin scintigraphy are generally recommended to assess for distant metastases, and a subsequent right colectomy to remove regional lymph nodes is performed if no distant metastases are observed. Whether right colectomy should be performed for tumors measuring <2 cm with features such as mesoappendiceal invasion or tumor origin at the appendiceal base, which in some series have suggested a poorer prognosis, remains uncertain. Additionally, tumors may arise in which neuroendocrine cells are admixed with mucin-producing cells or cells exhibiting features of frank adenocarcinoma. In such mixed neuroendocrine-adenocarcinoma tumors, sometimes termed “adenocarcinoids,” treatment recommendations are generally dictated by the more aggressive component of the tumor and align with typical recommendations for colorectal adenocarcinoma.

■ RECTAL NETS

With the increased use of screening colonoscopy, the diagnosis of rectal NET has also become more common. For unclear reasons, the incidence of rectal carcinoid tumors shows geographic variation. In European studies, they compose up to 14% of all NETs, while in some Asian series (Japan, China, Korea), they compose up to 90% of all NETs. The majority of rectal NETs are small, measuring <1 cm in diameter, and have well-differentiated histology. These tumors rarely

metastasize and can usually be safely removed endoscopically with subsequent endoscopic monitoring. In contrast, up to one-third of rectal NETs between 1 and 2 cm are associated with metastases, and those >2 cm, though uncommon, metastasize in >70% of patients. When identified early, these tumors generally require a surgical resection. In contrast to NETs of the appendix and small intestine, hormone secretion from rectal NETs, even when metastatic, is exceedingly rare.

CLINICAL PRESENTATION, DIAGNOSIS, AND EVALUATION OF PATIENTS WITH METASTATIC NEUROENDOCRINE TUMORS

While patients who undergo resection of localized NETs may be at risk of developing tumor recurrence or metastatic disease, postoperative treatment has not yet been shown to alter the risk of recurrence, and systemic adjuvant therapy is not recommended following resection of well-differentiated NETs, as it is for some other cancers. Whether adjuvant systemic therapy may be of benefit following resection of high-grade NETs is uncertain, and an approach utilizing platinum-based chemotherapy with or without external-beam radiation, analogous to that used in small-cell carcinoma, is sometimes considered.

The evaluation of patients with known or suspected metastatic disease generally includes both standard cross-sectional imaging such as CT or MRI and somatostatin scintigraphy. Somatostatin scintigraphy in this setting is based on the fact that >90% of NETs express somatostatin receptors. Gallium-68 (⁶⁸GA) dotate, a radioligand bound to a somatostatin analogue, can be used as a nuclear medicine tracer to perform PET scanning and is highly sensitive in detecting both primary NETs and metastases (Fig. 84-5). Because of the sensitivity of this approach, false-positive results can occur due to somatostatin receptor expression in other tissues. Physiologic uptake in the pancreatic uncinate process is common; uptake can also occur in the setting of sarcoidosis, in meningiomas, and in thyroid goiter or thyroiditis. Standard fluorodeoxyglucose (FDG) positron emission tomography (PET) scans are often negative in well-differentiated NET due to their low metabolic activity but can show uptake in higher-grade tumors; conversely, rates of somatostatin expression tend to be lower in higher-grade tumors, and ⁶⁸GA dotate scans may be negative in this setting.

The utility of blood-based tumor markers in NETs is controversial. The circulating tumor marker chromogranin A is commonly used as a screen for the presence of NETs and also to monitor for both

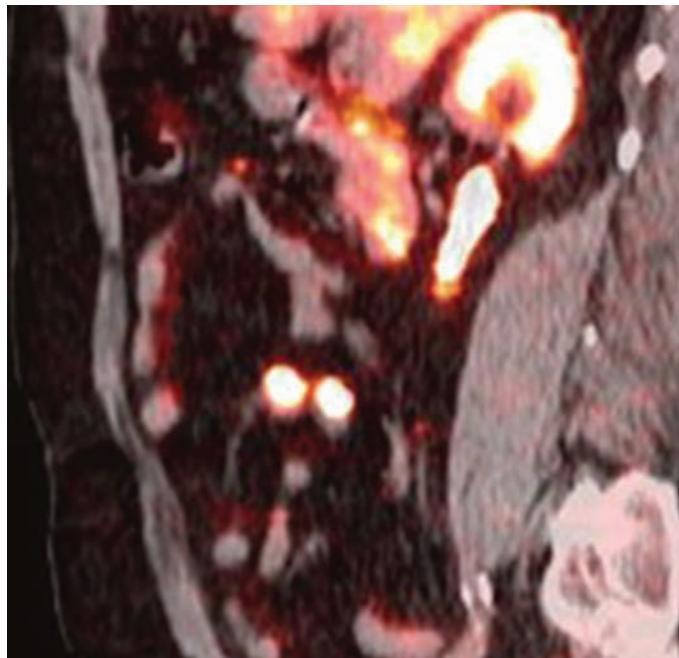


FIGURE 84-5 Gallium-68 Dotate PET scan demonstrating a small bowel neuroendocrine tumor and associated mesenteric mass. (Courtesy of Sara Meibom, MD, Boston Medical Center.)

recurrence and progression of disease in patients with known metastases. While chromogranin A is elevated in patients with metastatic NETs, it is neither particularly sensitive nor specific. A broad range of different assays for chromogranin A have also posed challenges in interpreting results in a standardized fashion. Chromogranin A is often elevated in a number of nonmalignant conditions, including in patients with impaired renal function and in patients who are taking PPIs. Elevated values of chromogranin A should be interpreted with caution in patients in whom a NET is being considered but in whom a diagnosis has not been established.

The overall survival durations for patients with metastatic NETs vary significantly, depending on both the primary location of the tumor and the histologic grade. Median survival durations for patients with well-differentiated NETs have markedly increased in recent years, likely reflecting both earlier diagnoses and improved treatments. For example, in early analyses of the SEER database, the median survival for patients with advanced pancreatic NETs was ~2 years; this had increased to 4 years in a more recent analysis. Similar increases were observed in patients with advanced small intestine NETs, where the median survival for patients with well-differentiated small intestine NETs exceeds 5 years. The sometimes prolonged survival of patients with NETs can sometimes make it challenging to determine at what point to initiate treatment. In patients with symptoms of hormone secretion, decisions to initiate therapy are straightforward. In asymptomatic patients, on the other hand, observation off treatment can sometimes be appropriate. Nevertheless, the natural course of NETs is ultimately to progress, and if treatment is not initiated, close monitoring is essential to ensure patients maximize access to available treatment options over the course of their disease.

MANAGEMENT OF SYMPTOMS OF HORMONE HYPERSECRETION AND THE CARCINOID SYNDROME

Patients with advanced NETs may in some cases experience more symptoms from hormone hypersecretion than from tumor bulk. The management of hormonal symptoms associated with pancreatic NETs depends on the hormone being secreted (see above). Patients with GI NETs, particularly those with small intestine or appendiceal primaries, may develop the carcinoid syndrome. Flushing and diarrhea are the two most common symptoms associated with carcinoid syndrome. The characteristic flush is of sudden onset; it is a deep red or violaceous erythema of the upper body, especially the neck and face, often associated with a feeling of warmth. Flushes may be precipitated by stress, alcohol, exercise, and certain foods such as cheese. Flushing episodes initially are brief, lasting 2–5 min, though later in the disease course, they may last hours. The diarrhea associated with carcinoid syndrome may or may not be associated with flushing and is described as watery in nature. Diarrhea can be profound, sometimes occurring in excess of 10 times daily and is one of the symptoms that most significantly interferes with activities of daily living. Less common manifestations of the carcinoid syndrome include wheezing or asthma-like symptoms. Impaired cognitive function has also been described in particularly advanced cases.

The main secretory product implicated in the carcinoid syndrome is serotonin (5-HT). Serotonin is synthesized from tryptophan by the enzyme tryptophan hydroxylase (Fig. 84-6). Up to 50% of dietary tryptophan can be used in this synthetic pathway by tumor cells, resulting in inadequate supplies for conversion to niacin; hence, some patients develop symptoms of niacin deficiency and pellagra-like lesions. Serotonin has numerous biologic effects, including the stimulation of intestinal secretion, increasing intestinal motility, and the stimulation of fibroblast growth. Other secreted products contributing to carcinoid syndrome symptoms are thought to include histamines and tachykinins, including substance P.

■ DIAGNOSIS AND TREATMENT OF THE CARCINOID SYNDROME

While the carcinoid syndrome can develop in patients with NETs from almost any site, it is most commonly associated with appendiceal or

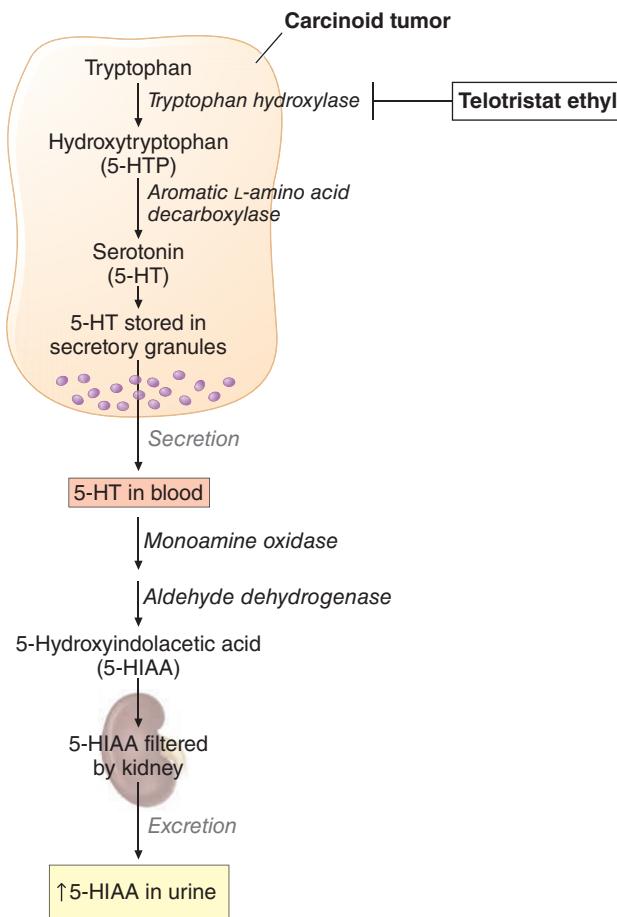


FIGURE 84-6 Serotonin synthesis and secretion in neuroendocrine tumors. Tryptophan is converted to hydroxytryptophan by tryptophan hydroxylase within the tumor cell and, subsequently, to serotonin (5-HT). Serotonin is subsequently converted to 5-hydroxyindole acetic acid (5-HIAA), which can be measured in a 24-h urine collection and can facilitate the diagnosis of carcinoid syndrome. Telotristat ethyl inhibits tryptophan hydroxylase and can be used as a treatment for carcinoid syndrome.

small intestine NETs. In these patients, the syndrome usually develops only after the development of hepatic metastases or retroperitoneal lesions, allowing entry of serotonin and other vasoactive substances into the systemic circulation. While serotonin levels can be measured in plasma, such measurements are frequently highly variable. Evidence of excess serotonin secretion can be more reliably confirmed by measuring levels of the serotonin metabolite 5-hydroxyindole acetic acid (5-HIAA), commonly using a 24-h urine collection. Urine collections can be challenging, and false-positive elevations may occur if the patient is eating serotonin-rich foods (e.g., salmon, eggs). As a result, elevated levels of 5-HIAA are suggestive but not diagnostic of the carcinoid syndrome. Patients with NETs may also experience symptoms of carcinoid syndrome related to other secreted products, including histamine, absent evidence of serotonin secretion. Conversely, patients without NETs may also describe symptoms analogous to carcinoid syndrome but due to other causes. The symptoms associated with systemic mastocytosis, in particular, can be easily confused with carcinoid syndrome.

The symptoms of carcinoid syndrome, including diarrhea, are generally refractory to standard antidiarrheals or other traditional medications but can often be well controlled with somatostatin analogues (Fig. 84-7). Approximately 90% of NETs express somatostatin receptors. The presence of somatostatin receptors on NETs can be easily confirmed with uptake on somatostatin scintigraphy such as ^{68}Ga dotate PET scan; uptake on somatostatin scintigraphy is predictive of response to treatment with somatostatin analogues. Somatostatin is a 14-amino-acid peptide that inhibits the secretion of a broad range of hormones. Due to its short half-life, administration of somatostatin

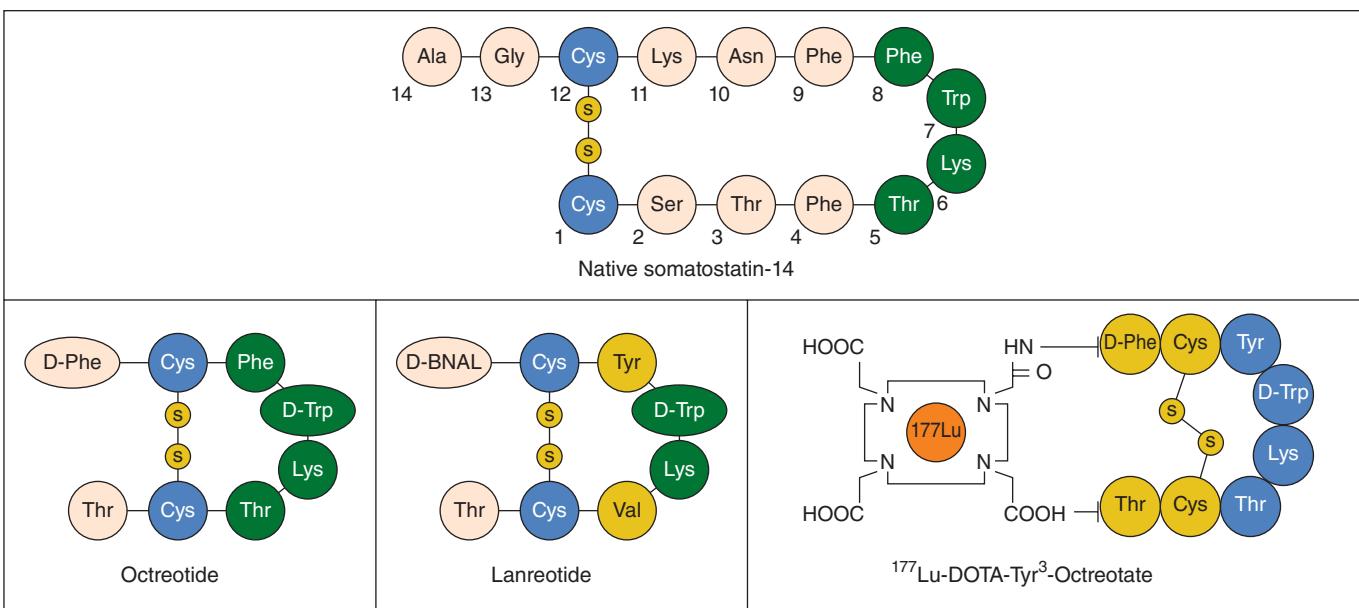


FIGURE 84-7 Somatostatin analogues. Commonly used somatostatin analogues include octreotide and lanreotide, which mirror the molecular structure of human somatostatin and bind to somatostatin receptors on neuroendocrine tumors. Somatostatin analogues inhibit tumoral hormone secretion and also have an antiproliferative effect. Radiolabeled somatostatin analogues such as ^{177}Lu -DOTA-octreotate, shown in the figure, share a similar molecular structure and are used therapeutically.

itself is not therapeutically practical. Longer-acting somatostatin analogues, including octreotide and lanreotide, share an 8-amino-acid binding domain with naturally occurring somatostatin and bind primarily to somatostatin receptor subtypes 2 and 5. Both have been shown to be effective in the treatment of carcinoid syndrome.

In an initial study, treatment of patients with octreotide 150 µg subcutaneously three times daily controlled symptoms of flushing and diarrhea in 88% of patients. A depot preparation (octreotide long-acting release [LAR]) that can be administered monthly has largely eliminated the need for daily octreotide injections and is now considered a standard approach for symptomatic treatment of advanced NETs. Lanreotide, another long-acting somatostatin analogue that is also administered monthly, appears to have similar clinical efficacy to octreotide in the treatment of metastatic NETs and the carcinoid syndrome. As described further below, both octreotide and lanreotide also share the ability to slow tumor growth, providing an additional benefit to patients.

Somatostatin analogue side effects are generally mild. Mild nausea, abdominal discomfort, bloating, and loose stools occur in up to one-third of patients during the first month or two of treatment but usually subsequently subside. Patients with persistent symptoms of bloating or loose stools may be experiencing pancreatic insufficiency associated with use of somatostatin analogues; use of pancreatic enzyme supplements can ameliorate these symptoms. Mild glucose intolerance may also occur due to inhibition of insulin secretion. One of the more significant side effects associated with somatostatin analogues is impaired gallbladder contractility, resulting in delayed gallbladder emptying, and long-term administration of somatostatin analogues has been associated with an increased risk of cholelithiasis. For this reason, patients with advanced NETs in whom surgery is planned and for whom somatostatin analogue therapy is being considered should generally also undergo prophylactic cholecystectomy.

Over time, for reasons that remain uncertain, patients receiving somatostatin analogues for symptoms of hormone secretion may become refractory to treatment. Not uncommonly, such patients experience symptom exacerbation toward the final week of each treatment cycle. Such patients may benefit from an increased frequency of administration (i.e., every 3 weeks) or use of additional short-acting octreotide for breakthrough symptoms.

The association between high levels of circulating serotonin and symptoms of the carcinoid syndrome has also led to a strategy aiming to directly inhibit serotonin synthesis (Fig. 84-6). This approach was

first undertaken in the late 1960s with the drug para-chlorophenylalanine, which was reported to reduce symptoms of carcinoid syndrome but also caused significant central nervous system (CNS) side effects. Telotristat ethyl, a tryptophan hydroxylase inhibitor with minimal CNS penetration, was evaluated in a randomized trial that enrolled 135 patients with persistent carcinoid syndrome-related diarrhea while receiving somatostatin analogues. Treatment with telotristat ethyl was associated with a reduction in bowel movement frequency as well as significant decreases in urinary 5-HIAA compared to placebo. Thus, telotristat is a treatment option for patients with carcinoid syndrome who have persistent diarrhea despite treatment with somatostatin analogues.

CARCINOID CRISIS

Carcinoid crisis has been described in the setting of tumor manipulation during surgery and, less commonly, after other interventions such as hepatic artery embolization or radionuclide therapy. It may also occur as a result of exogenous administration of epinephrine or during induction of anesthesia. It is most common in patients who already have significant symptoms of carcinoid syndrome and is thought to be caused by a sudden release of biologically active compounds from the tumor. Carcinoid crisis can be life-threatening and can manifest as either profound hypotension or hypertension. Prospective studies on the prevention and management of carcinoid crisis are limited; however, somatostatin analogues should be readily available during surgical procedures, and in some cases, continuous prophylactic intravenous administration of somatostatin analogues has been utilized as a way to mitigate risk.

CARCINOID HEART DISEASE

Carcinoid heart disease occurs in approximately two-thirds of patients with the carcinoid syndrome. Carcinoid heart lesions are characterized by plaque-like, fibrous endocardial thickening that classically involves the right side of the heart and often causes retraction and fixation of the leaflets of the tricuspid and pulmonary valves (Fig. 84-8). The fibrosis in carcinoid heart disease is thought to be directly related to exposure of heart valve fibroblasts to high circulating levels of serotonin. Lesions similar to those observed in carcinoid heart disease were observed in patients receiving fenfluramine, a drug also known to increase serotonin signaling, as well as in patients receiving ergot-containing dopamine receptor agonists for Parkinson's disease. Metabolites of fenfluramine, as well as the dopamine receptor agonists, have high

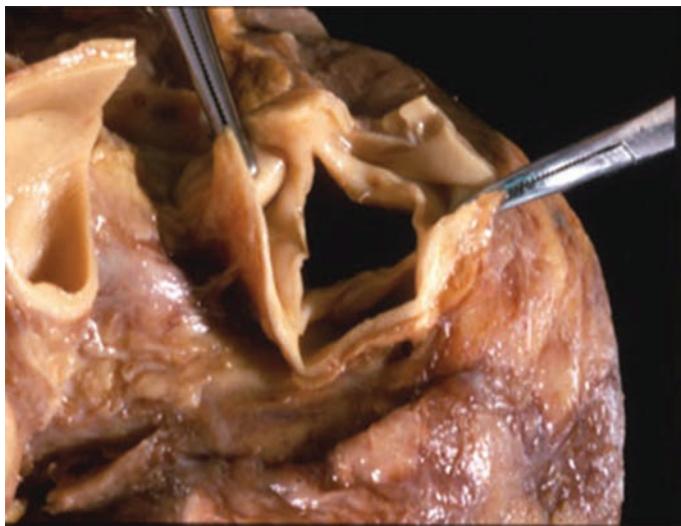


FIGURE 84-8 Carcinoid heart disease. Fibrosis secondary to elevated levels of circulating serotonin classically involves the tricuspid valve, resulting in valve retraction and tricuspid regurgitation.

affinity for serotonin receptor subtype 5-HT_{2B} receptors, whose activation is known to cause fibroblast mitogenesis and which are normally expressed in heart valve fibroblasts. These observations support the hypothesis that serotonin overproduction in patients with carcinoid syndrome mediates the valvular changes by activating 5-HT_{2B} receptors in the endocardium.

Tricuspid regurgitation is a nearly universal feature of carcinoid heart disease; tricuspid stenosis, pulmonary regurgitation, and pulmonary stenosis may also occur. Left-sided heart disease occurs in <10% of patients and has been associated with the presence of a patent foramen ovale. The preponderance of lesions in the right heart is related directly to the fact that serotonin is secreted by liver metastases or retroperitoneal tumor deposits into the venous circulation and subsequently into the right atrium and right ventricle. The lower incidence of heart disease in the left heart is postulated to be due to the fact that serotonin is metabolized in the pulmonary vasculature before entering the left atrium and ventricle. Among patients with carcinoid syndrome, patients with heart disease exhibit higher levels of serum serotonin and urinary 5-HIAA excretion than patients without heart disease. Treatment with somatostatin analogues resulting in decreased serotonin secretion does not result in regression of cardiac lesions. Reduction of serotonin levels as a result of treatment with somatostatin analogues or with the tryptophan hydroxylase inhibitor telotristat ethyl seems likely to slow progression of carcinoid heart disease but has not been formally evaluated in clinical trials.

Right-sided heart failure in patients with carcinoid heart disease may lead to significant morbidity and mortality. The development of multiple new treatments to improve overall disease control in patients with advanced NETs has led to increased interest in valvular replacement, which may result in significant clinical benefit in appropriately selected patients with carcinoid heart disease. The appropriate timing of valve replacement in such patients can be challenging given the competing desires to perform surgery before the onset of severe right-sided heart failure, which can increase surgical morbidity, and the need to achieve adequate overall tumor control. However, advanced and less invasive techniques, including catheter-based valve replacement, have made valve replacement an increasingly attractive option for patients with this condition.

HEPATIC-DIRECTED THERAPY FOR METASTATIC NETS

The liver is one of the most common sites for metastases in patients with NETs and, in some cases, is the only site of metastatic disease. Hepatic-directed therapies can often be effective as a means of controlling, if not eliminating, metastases, particularly in patients who have more indolent tumors with well-differentiated histology. Common

approaches for such patients include surgical resection, ablation or embolization, and orthotopic liver transplantation.

For patients with limited hepatic disease whose tumors have well-differentiated histology, surgical resection is generally considered the preferable option. While data are limited to retrospective series with the consequent risk of selection bias, long-term survival durations and symptomatic improvements reported in select populations of patients undergoing hepatic resection of neuroendocrine liver metastases compare favorably with outcomes associated with other management approaches, and 5-year survival rates approach 90% in some series. In patients in whom anatomy precludes resection or in whom a greater number of lesions are present, radiofrequency ablation or cryoablation can also be used, either as a primary treatment modality or as an adjunct to surgical resection. While ablation is considered to be less morbid than hepatic resection, it is generally utilized only in smaller tumors so that zones of ablation are limited.

In most cases, however, liver metastases are large, multiple, and involve both lobes of the liver. In such cases, the benefit of surgical resection and ablation is limited. Hepatic arterial embolization can be considered in these cases, assuming that extrahepatic disease remains relatively limited and that clinical benefit can be achieved by reducing hepatic tumor bulk. Hepatic artery embolization is based on the principle that tumors in the liver derive most of their blood supply from the hepatic artery, whereas healthy hepatocytes derive most of their blood supply from the portal vein. Multiple different embolization techniques have been explored, ranging from the simple infusion of gel foam powder into the hepatic artery (bland embolization) to the administration of chemotherapy or chemotherapy-eluting beads into the hepatic artery (chemoembolization) or the intra-arterial administration of radioisotope-tagged microspheres (radioembolization). Limited data suggest an optimal approach to embolization, and few studies have compared these approaches directly. Tumor response rates with all of these approaches generally exceed 50%. Specific approaches are therefore often tailored to the patient, taking into account tumor location, overall tumor burden, and comorbidities. Bland embolization, for example, may be associated with less morbidity, whereas chemoembolization or radioembolization may result in longer durations of response.

The role of orthotopic liver transplantation for the treatment of NETs remains uncertain. Data from available institutional series suggest that a small number of highly selected patients may achieve long-term survival. However, 5-year overall median survival durations in most series are ~50%, and the majority of patients undergoing hepatic transplantation develop tumor recurrence. Additionally, the widespread utility of hepatic transplantation is limited by organ availability. Decisions regarding proceeding with transplantation in patients with advanced NETs are therefore highly individualized.

SYSTEMIC TREATMENT TO CONTROL TUMOR GROWTH

While hepatic-directed therapies can be effective in the management of patients with liver-predominant disease, a majority of patients will either present with or ultimately develop more widespread metastases. A number of systemic treatment options have been developed and can be effective in treating such patients. These options include treatment with traditional somatostatin analogues, peptide receptor radioligand therapy, traditional cytotoxic chemotherapy, and an increasing array of molecularly targeted therapies targeting the mTOR or vascular endothelial growth factor (VEGF) pathways (Table 84-3). The choice and sequence of therapy depend in part on the type of tumor, the extent of disease, and patient symptoms and comorbidities.

SOMATOSTATIN ANALOGUES

While somatostatin analogues were originally developed as a treatment to reduce hormone secretion in NETs, they are also effective in slowing tumor growth. The biologic mechanisms underlying this effect remain uncertain, but clinical studies have been definitive. The first of these studies, the PROMID study, randomized patients with metastatic small-intestinal NET to receive either octreotide LAR at a

TABLE 84-3 Selected Randomized Trials of Therapeutic Agents for the Treatment of Advanced Neuroendocrine Tumors (NETs)

TUMOR TYPE	NUMBER OF PATIENTS	PROGRESSION-FREE SURVIVAL
Pancreatic and extrapancreatic NET		
Lanreotide vs placebo (CLARINET)	204	65 vs 33% at 2 years ($p < .001$)
Pancreatic NET		
Everolimus vs placebo (RADIANT 3)	410	11 months vs 4.6 months ($p < .001$)
Sunitinib vs placebo	171	11.4 months vs 5.5 months ($p < .001$)
Surufatinib vs placebo	264	10.9 months vs 3.7 months ($p = .001$)
Temozolomide/capecitabine vs temozolomide	144	22.7 months vs 14.4 months ($p = .021$)
Extrapancreatic NET		
Octreotide vs placebo (PROMID)	85	14.3 months vs 6 months ^a
Everolimus + octreotide vs octreotide (RADIANT 2)	429	16.4 months vs 11.3 months
Everolimus vs placebo (RADIANT 4)	302	11 months vs 3.9 months
Surufatinib vs placebo	198	9.2 months vs 3.8 months ($p < .0001$)
Pazopanib vs placebo	171	11.6 months vs 8.5 months ($p < .0005$)
177-Lutetium dotatate vs octreotide (NETTER 1)	230	65.2 vs 10.8% at 20 months ($p < .001$)

^aTime to tumor progression.

dose of 30 mg monthly or placebo. The median time to tumor progression in patients receiving octreotide was 14 months compared to only 6 months for patients receiving placebo. Because the study was limited to patients with small-intestinal NET, the generalizability of these results to patients with NETs of other origins, including pancreatic NET, was initially uncertain. This question was ultimately addressed by the phase 3 CLARINET trial, which compared lanreotide, a somatostatin analogue that is similar to octreotide in its somatostatin receptor-binding affinities, to placebo in 204 patients with a range of advanced well- or moderately differentiated gastroenteropancreatic NETs. Progression-free survival duration at 2 years was 65% in patients receiving lanreotide and 33% in patients receiving placebo, a difference that was statistically significant. One unusual aspect of the PROMID and CLARINET studies is the difference in progression-free survival durations in the placebo arms of the studies, which has been attributed to differences in patient selection. Either octreotide or lanreotide is currently considered an acceptable option for control of tumor growth in patients with advanced NETs.

Whether treatment with somatostatin analogues also increases overall survival in patients with advanced NETs has not been demonstrated, although a correlation between progression-free survival and overall survival in patients with advanced NETs treated with single-agent somatostatin analogue therapy has been shown. The timing of initiation of somatostatin analogues in patients with advanced NETs remains uncertain. The variable clinical course of NETs means that tumors can remain indolent for years even without treatment. For patients with asymptomatic, small-volume disease, observation alone may be an appropriate initial option. However, for patients with a larger disease burden, evidence of disease progression, or symptomatic disease, somatostatin analogues are generally used as an initial systemic treatment due to their ease of use and tolerability.

■ PEPTIDE RECEPTOR RADOLIGAND THERAPY

Peptide receptor radioligand therapy employs the systemic administration of radiolabeled somatostatin analogues and is a treatment option for patients who require more aggressive treatment due to progression

on traditional somatostatin analogues or other therapies (Fig. 84-7). Peptide receptor radioligand therapy may also be considered as an initial treatment in patients with significant symptoms or tumor burden. With this approach, a radioligand is coupled to a somatostatin analogue, using the somatostatin analogue to target the tumor. When bound to the tumor cell, the radioligand is then internalized, resulting in cell death. Due to its mechanism of action, peptide receptor radioligand therapy is only considered in patients whose tumors demonstrate uptake on somatostatin scintigraphy.

Several different radioligands have been evaluated, the most successful of which have been yttrium (⁹⁰Y) and lutetium (¹⁷⁷Lu). These two ligands differ from one another in terms of their particle energy and tissue penetration; of the two, ⁹⁰Y-DOTA-TOC emits higher-energy β particles and has deeper tissue penetration. ⁹⁰Y-DOTA-TOC (⁹⁰Y-dotatoc) has been evaluated in numerous series with overall tumor responses reported in approximately one-third of patients. Enthusiasm for this approach, however, has been tempered due to concerns about side effects including both renal and hematologic toxicity.

¹⁷⁷Lu-DOTA-octreotide emits both β particles and lower-energy γ particles and, in most studies, has been associated with less toxicity than ⁹⁰Y-DOTA-TOC. Initial single-center studies with ¹⁷⁷Lu-DOTA-octreotide showed promising antitumor activity, and based on these studies, a randomized trial of ¹⁷⁷Lu-dotatate in midgut GI NETs was undertaken. In this study (NETTER-1), 229 patients with inoperable, somatostatin receptor-positive midgut NETs were randomly assigned to receive either four doses of ¹⁷⁷Lu-dotatate administered intravenously every 8 weeks or treatment with high-dose octreotide LAR (60 mg) every 4 weeks. Treatment with ¹⁷⁷Lu-dotatate was associated with objective tumor responses in 18% of patients and also was associated with a significant improvement in progression-free survival: progression-free survival at month 20 was 10.8% for octreotide LAR alone and 65.2% in the ¹⁷⁷Lu-dotatate group. Subsequent analyses have also suggested improved overall survival associated with ¹⁷⁷Lu-dotatate treatment, as well as improvements in quality of life across a number of parameters, including global health status, overall physical functioning, fatigue, pain, and diarrhea.

The renal clearance of radiopeptides, including ¹⁷⁷Lu-DOTA-octreotide, poses a risk of renal toxicity. The renal toxicity can be mitigated with the coadministration of intravenous amino acids during treatment. The most common adverse event among patients receiving ¹⁷⁷Lu-dotatate in the NETTER-1 study was nausea, most likely related to the amino acid infusions rather than to the radioisotope itself. Mild thrombocytopenia and leukopenia were also reported.

One limitation of the NETTER-1 study was its restriction to patients with advanced small intestine NETs. However, longer-term safety data as well as data supporting the efficacy of ¹⁷⁷Lu-dotatate in a broader range of gastroenteropancreatic NETs are available from large institutional series that include >1000 patients. Long-term toxicities from these series have included rare cases of acute leukemia and myelodysplastic syndrome, presumably associated with radiation exposure. Nevertheless, these studies generally support both the efficacy and safety of ¹⁷⁷Lu-dotatate as a treatment for patients with a range of somatostatin receptor-positive NETs.

■ ALKYLATING AGENTS

While the efficacy of traditional cytotoxic chemotherapy appears to be minimal in most extrapancreatic GI NETs, alkylating agents have a clear role in the treatment of advanced pancreatic NETs. Streptozocin-based combination therapy was historically used as treatment standard in such patients but has largely fallen out of favor due to both toxicity concerns and a cumbersome administration schedule. Temozolomide is an orally administered alkylating agent that has largely replaced streptozocin as a backbone in combination regimens used for the treatment of pancreatic NETs.

Initial studies evaluating temozolomide in combination with a range of different agents showed that temozolomide-based combination therapy was associated with tumor responses in 24–70% of patients. One of the most active combination regimens appeared to be temozolomide and capecitabine. This combination was subsequently compared to

temozolomide alone in a prospective randomized study undertaken by the Eastern Cooperative Oncology Group that enrolled 144 patients with advanced pancreatic NETs. The overall response rates in the two arms were relatively similar; 33% of patients who received the combination of temozolomide and capecitabine experienced objective tumor responses as compared to 28% of the patients who received temozolomide as a single agent. However, progression-free survival was significantly longer in the combination arm (22.7 vs 14.4 months). Based on these results, the combination of temozolomide and capecitabine is now the preferred chemotherapy combination for advanced pancreatic NETs.

The reason that some pancreatic NETs respond to alkylating agents while others do not is uncertain. In patients with glioblastoma, methylation of the promoter region for methylguanine DNA methyltransferase (MGMT) is associated with decreased MGMT protein expression and is highly associated with temozolomide responsiveness. MGMT is an enzyme that is responsible for repairing DNA damage induced by alkylating agents. Reduced levels of MGMT presumably impair the ability of tumor cells to repair their DNA in response to treatment and enhance the cytotoxicity of temozolomide. Several retrospective studies have suggested that lack of MGMT expression in pancreatic NET may be associated with responsiveness to temozolomide-based therapy; however, this finding has not yet been prospectively validated.

SMALL-MOLECULE TYROSINE KINASE INHIBITORS

The highly vascular nature of NETs combined with observations in preclinical models that disruption of signaling pathways associated with VEGF inhibits neuroendocrine cell growth prompted a number of clinical trials evaluating therapeutic agents that inhibit the VEGF pathway in both pancreatic and extrapancreatic NETs. The VEGF pathway is activated through the binding of VEGF to its cell surface receptor, which initiates an intracellular signaling cascade that promotes angiogenesis as well as cell growth, proliferation, and survival. Clinical trials of VEGF pathway inhibitors in NETs have included a number of small-molecule tyrosine kinase inhibitors that, while they differ to some extent in specificity, all have in common the property targeting VEGFR2, the receptor isoform most strongly implicated in promoting angiogenesis.

Sunitinib, a multitargeted tyrosine kinase inhibitor that inhibits a range of growth factor receptors including VEGFR2, was one of the first agents in this class found to have activity in pancreatic NETs. In an initial phase 2 trial, sunitinib was administered to 109 patients with either pancreatic or extrapancreatic NET. Of 61 patients with pancreatic NET enrolled in the study, 11 had evidence of an objective tumor response. Based on these observations, sunitinib was evaluated in an international, randomized trial in which continuous administration of sunitinib (37.5 mg daily) was compared with placebo in 171 patients with advanced, progressive pancreatic NET. The median progression-free survival was significantly longer in patients treated with sunitinib compared with patients treated with placebo (11.4 vs 5.5 months). Common side effects associated with sunitinib included hypertension, proteinuria, and fatigue.

A second VEGFR-targeted tyrosine kinase inhibitor, surufatinib, has been evaluated in a randomized trial in which 264 patients with advanced pancreatic NETs from 21 centers in China were randomized to receive either surufatinib, administered at a dose of 300 mg daily, or placebo. Patients receiving surufatinib experienced a median progression-free survival duration of 10.9 months, as compared to 3.7 months in patients receiving placebo, closely mirroring the results of the earlier sunitinib study. Other small-molecular tyrosine kinase inhibitors have been evaluated in smaller, single-arm studies and have shown activity in pancreatic NETs, including sorafenib, cabozantinib, pazopanib, and axitinib.

Small-molecule tyrosine kinase inhibitors targeting the VEGF pathway have also been evaluated in patients with advanced nonpancreatic GI NET. In most of these studies, objective tumor response rates are lower than those seen in pancreatic NET, though many of these initial studies also revealed low rates of tumor progression and encouraging

progression-free survival durations, suggesting that these agents had antitumor activity. Pazopanib was compared to placebo in a randomized study undertaken by the ALLIANCE cooperative group, which enrolled 171 patients with nonpancreatic NETs. Patients treated with pazopanib in this study had a superior progression-free survival compared to those who received placebo (11.6 vs 8.5 months), a difference that was statistically significant. Surufatinib was used in a randomized study of 198 patients with extrapancreatic NETs; the median progression-free survival was 9.2 months in patients receiving surufatinib and 3.8 months in those receiving placebo, a statistically significant difference. These studies suggest that VEGF-targeted tyrosine kinase inhibitors have antitumor activity in extrapancreatic and pancreatic NETs.

mTOR INHIBITORS

mTOR is an intracellular protein kinase that has been implicated in the regulation of a number of processes regulating cell growth in both normal and malignant cells. It functions as a downstream component of the PI3-AKT-mTOR pathway. This pathway is negatively regulated by the tuberous sclerosis complex, comprising TSC1 and TSC2. An association between the development of pancreatic NETs and inherited mutations in TSC2 in patients with tuberous sclerosis complex was a contributing factor to initial interest in exploring mTOR inhibition as a therapeutic approach in this setting.

Following initial evidence of antitumor activity associated with everolimus (10 mg daily) in an international, multicenter, phase 2 trial of 160 patients, everolimus monotherapy (10 mg daily) was compared with best supportive care alone in the RADIANT-3 trial that enrolled 410 patients with advanced progressing pancreatic NET. While overall objective responses were uncommon, treatment with everolimus was associated with a significant prolongation in median progression-free survival (11.0 vs 4.6 months) compared to placebo, supporting its use as a standard treatment to control tumor growth in patients with advanced pancreatic NET. Common toxicities associated with everolimus are generally mild and can include stomatitis and rash; a more severe but less common side effect is pneumonitis.

Everolimus was also associated with promising activity in early phase 2 studies enrolling patients extrapancreatic NET. The first large randomized study evaluating everolimus was the RADIANT 2 trial; 429 patients with advanced GI NETs were randomly assigned to receive octreotide LAR (30 mg intramuscularly every 28 days) with or without everolimus (10 mg daily). Treatment with everolimus in this study was associated with an improvement in median progression-free survival (16.4 vs 11.3 months), but the difference in this study was of only borderline statistical significance. A second study, the RADIANT 4 study, enrolled 302 patients with advanced NETs of either GI (excluding pancreatic) or lung origin, randomizing them to receive either everolimus or placebo. In this study, treatment with octreotide was not required. As in the RADIANT 3 study, objective tumor responses were uncommon; however, median progression-free survival in patients who received everolimus was significantly longer than in those who received placebo (11 vs 3.9 months). Based on the results of this study, everolimus is considered a standard treatment for control of tumor growth in extrapancreatic NETs.

OTHER SYSTEMIC TREATMENTS FOR CONTROL OF TUMOR GROWTH

Interferon α has been used as a treatment for advanced NETs for several decades. With the development of newer approaches, its routine use has diminished. The use of interferon α was based primarily on observations in large, retrospective series where low-dose interferon α was reported to both reduce symptoms of hormonal hypersecretion and slow tumor progression. Interferon can be myelosuppressive, requiring dose titration, and in some patients can induce both fatigue and depression. Antitumor activity has also been reported with oxaliplatin-based chemotherapy regimens. A combined analysis of two phase 2 trials examining oxaliplatin-fluoropyrimidine chemotherapy plus bevacizumab in advanced NET suggested antitumor activity for these regimens; the benefit appeared to be greatest in patients with intermediate-grade rather than low-grade tumors.

■ SYSTEMIC THERAPY FOR HIGH-GRADE NEUROENDOCRINE CARCINOMA

High-grade NETs are relatively uncommon; their clinical behavior is fundamentally different from well-differentiated NETs in that these tumors pursue an aggressive clinical course. Systemic chemotherapy for advanced high-grade neuroendocrine carcinoma has historically followed a paradigm analogous to that used for small-cell carcinoma of the lung, with combinations of either cisplatin or carboplatin administered together with etoposide generally considered the preferred first-line approach. One of the most important elements in determining the optimal chemotherapeutic approach is assessing the Ki-67 proliferative index. A large retrospective series that evaluated 252 patients with high-grade neuroendocrine carcinoma found that the activity of platinum-based therapy was greatest in patients who had a Ki-67 proliferative index of 55% or higher; in these patients, the overall tumor response rate was 42%. In contrast, the overall response rate in patients in whom the Ki-67 proliferative index was <55% was only 15%. As in small-cell carcinoma of the lung, immune checkpoint inhibitors also appear to have some activity in high-grade neuroendocrine carcinoma. While minimal activity has been noted in well-differentiated NETs, a combination of ipilimumab and nivolumab was associated with an overall tumor response rate of 42% in an initial phase 2 trial enrolling 19 patients with high-grade neuroendocrine carcinoma.

ACKNOWLEDGEMENT

Dr. Robert Jensen contributed this chapter in previous editions, and some material from his chapter is retained here.

■ FURTHER READING

- CAPLIN ME et al: Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *N Engl J Med* 371:224, 2014.
- DASARI A et al: Trends in the incidence, prevalence, and survival outcomes for patients with neuroendocrine tumors in the United States. *JAMA Oncol* 3:1336, 2017.
- KULKE MH et al: Telotristat ethyl, a tryptophan hydroxylase inhibitor for the treatment of carcinoid syndrome. *J Clin Oncol* 35:14, 2017.
- RAYMOND E et al: Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med* 364:501, 2011.
- RINDI G et al: A common classification framework for neuroendocrine neoplasms: An International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. *Mod Pathol* 31:1770, 2018.
- SCARPA A et al: Whole genome landscape of pancreatic neuroendocrine tumors. *Nature* 543:65, 2017.
- STROSBERG J et al: Phase 3 trial of 177 Lu-dotatate for midgut neuroendocrine tumors. *N Engl J Med* 376:125, 2017.
- YAO JC et al: Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med* 364:514, 2011.

■ EPIDEMIOLOGY

The incidence of cancers of the kidney and renal pelvis rose for three decades, reached a plateau of approximately 64,000 cases annually in the United States between 2012 and 2018, but has since increased to approximately 76,000 cases annually, resulting in close to 14,000 deaths per year. It is the eighth most common cancer overall in the United States, the sixth most common in males, and the eighth most common in females; the male-to-female ratio is 2:1. Although this malignancy may be diagnosed at any age, it is uncommon in those under 45 years, and incidence peaks between the ages of 55 and 75 years. Many factors have been investigated as possible contributing causes; associations include cigarette smoking, obesity, and hypertension. Risk is also increased for patients with polycystic kidney disease that has been complicated by chronic renal failure.

Most cases of renal cell carcinoma (RCC) are sporadic, although familial forms have been reported (**Table 85-1**). One well established example includes clear cell RCC arising in the context of von Hippel-Lindau (VHL) syndrome, an autosomal dominant disorder. Genetic studies identified the *VHL* gene on the short arm of chromosome 3. Individuals with VHL syndrome have an estimated lifetime risk of developing clear cell RCC of around 70%. Other *VHL*-associated neoplasms include retinal hemangioma, hemangioblastoma of the spinal cord and cerebellum, pheochromocytoma, and neuroendocrine tumors and cysts. Birt-Hogg-Dubé syndrome is a rare human autosomal dominant genetic disorder characterized by fibrofolliculomas (benign tumors arising in hair follicles), pulmonary cysts, and renal cell carcinomas of varying histologies, most commonly the chromophobe type, occurring in about a third of patients. This disorder is associated with mutations in the *FLCN* gene, which codes for folliculin. Other hereditary syndromes are summarized in Table 85-1.

■ PATHOLOGY AND GENETICS

Renal cell malignancies represent a heterogeneous group of tumors with distinct histopathologic, genetic, and clinical features (**Table 85-2**). Categories include clear cell carcinoma (70% of cases), papillary tumors (10–15%), chromophobe tumors (≤5%), renal medullary carcinoma (<1%), translocation carcinoma (<5%), and other less common variants. Papillary tumors can be bilateral and multifocal. Chromophobe tumors tend to have a more indolent clinical course. Translocation-associated RCC, rare in adult patients, is the predominant histology in children. Renal medullary carcinoma is rare, very aggressive, and associated with sickle cell trait. Tumors that do not meet criteria for defined variants are generally referred to as “unclassified” with variable clinical courses.

Clear cell tumors, the predominant histology, are found in >80% of patients who develop metastases and arise from the epithelial cells of the proximal tubules. Loss of chromosome 3p is uniformly seen as the earliest event in the development of these cancers. This leads to loss of heterozygosity for a number of relevant 3p genes, including *VHL*, *PBRM1*, *BAP1*, and *SETD2*, which can be functionally lost through secondary events in the remaining allele. *VHL* encodes a tumor suppressor protein that is involved in regulating the transcription of vascular endothelial growth factor (VEGF) and a number of other effectors through ubiquitination of hypoxia-inducible factors (HIF). Inactivation of *VHL*, through upregulation of VEGF signaling, promotes tumor angiogenesis and growth, ultimately rendering clear cell RCC cells susceptible to antiangiogenesis therapy.

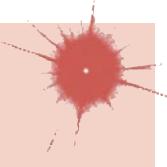
Large-scale sequencing efforts have helped elucidate recurrent patterns of genomic evolution that correlate with distinct clinical phenotypes, e.g., varying levels of aggressiveness or specific patterns of metastatic spread. For example, early loss of chromosome 9p appears to confer a high risk for early metastatic dissemination and correlates with poor cancer-specific survival.

A growing number of other RCC variants are well defined (see Table 85-2 for examples). For instance, up to 15% of RCCs are of the papillary subtype, with several subtypes that can be distinguished either by light microscopy or tumor genomics. For example, activating mutations in the *MET* oncogene or gain of chromosome 7 (where *MET* is located) are hallmark events of type 1 papillary RCC and considered

85

Renal Cell Carcinoma

Robert J. Motzer, Martin H. Voss



Renal cell carcinomas account for 90–95% of malignant neoplasms arising from the kidney. Notable features include frequent diagnosis without symptoms, resistance to cytotoxic agents, robust activity of angiogenesis-targeted agents, immune infiltration commonly rendering tumors susceptible to checkpoint-directed immunotherapy, and a variable clinical course for patients with metastatic disease, including anecdotal reports of spontaneous regression. Most of the remaining 5–10% of malignant neoplasms arising from the kidney are transitional cell carcinomas (urothelial carcinomas) originating in the lining of the renal pelvis. See **Chap. 86** for transitional cell carcinomas.

TABLE 85-1 Hereditary Renal Cell Tumors

SYNDROME	CHROMOSOME(S)	GENE	PROTEIN	KIDNEY TUMOR TYPE	ADDITIONAL CLINICAL FINDINGS
von Hippel-Lindau syndrome	3p25	VHL	von Hippel-Lindau protein	Clear cell	Hemangioblastoma of the retina and central nervous system; pheochromocytoma; pancreatic and renal cysts; neuroendocrine tumors
Hereditary papillary RCC	7p31	MET	MET	Papillary (type I)	Bilateral and multifocal kidney tumors
Hereditary leiomyomatosis and RCC (HLRCC syndrome)	1q42	FH	Fumarate hydratase	Papillary (non-type I)	Leiomyoma; uterine leiomyoma/leiomyosarcoma
Birt-Hogg-Dubé syndrome	17p11	FLCN	Folliculin	Chromophobe; oncocytoma	Facial fibrofolliculoma; pulmonary cysts
Tuberous sclerosis	9q34 16p13	TSC1 TSC2	Hamartin Tuberin	Angiomyolipomas; lymphangioleiomyomatosis; rare RCC with variety of histologic appearances	Angiofibroma, subungual fibroma; cardiac rhabdomyoma; adenomatous small intestine polyps; pulmonary and renal cysts; cortical tuber; subependymal giant cell astrocytomas
BAP1 tumor predisposition syndrome	3p21	BAP1	BAP1	Clear cell	Atypical Spitz tumors; uveal melanoma; cutaneous melanoma; basal cell carcinoma; malignant mesothelioma

Abbreviation: RCC, renal cell carcinoma.

actionable via targeted MET inhibitors. Tumors of the less common chromophobe subtype originate from the distal nephron. They are in part driven by changes in mitochondrial gene function and typically characterized by aneuploidy with common loss of an entire chromosome copy for chromosomes 1, 2, 6, 10, 13, and 17.

CLINICAL PRESENTATION

Presenting signs and symptoms may include hematuria, flank or abdominal pain, and a palpable mass. Other symptoms are fever, weight loss, anemia, and a varicocele. Tumors are, however, commonly detected as an incidental finding on a radiograph. Widespread use of radiologic cross-sectional imaging (computed tomography [CT], magnetic resonance imaging [MRI]) contributes to earlier detection of renal masses during evaluation for other medical conditions. The increasing number of incidentally discovered low-stage tumors has contributed to an improved 5-year survival for patients with RCC and increased use of nephron-sparing surgery (partial nephrectomy). A spectrum of paraneoplastic syndromes has been associated with these malignancies, including erythrocytosis, hypercalcemia, nonmetastatic hepatic dysfunction (Stauffer's syndrome), and acquired dysfibrinogenemia. Erythrocytosis is noted at presentation in only about 3% of patients. Anemia, commonly a sign of more advanced disease, is more common. Kidney cancer was called the "internist's tumor" since it was often discovered from the initial presentation of a paraneoplastic syndrome. This was more common before the era of modern imaging, as was initial presentation by the classic triad of hematuria, flank pain, and a palpable abdominal mass.

The standard evaluation of patients with suspected renal tumors includes a CT scan of the abdomen and pelvis, chest radiograph, and urine analysis. If metastatic disease is suspected from the chest radiograph, a CT of the chest is warranted. MRI is useful in evaluating the

inferior vena cava in cases of suspected tumor involvement or invasion by thrombus, or when intravenous contrast administration given with CT is prohibited by impaired renal function. In clinical practice, any solid renal masses should be considered malignant until proven otherwise; a definitive diagnosis is required. If no metastases are demonstrated, surgery is indicated, even if the renal vein or inferior vena cava is invaded. In small tumors (particularly those of clear cell variant) the risk of impending metastatic spread is lower and surgery can potentially be delayed. In that setting, a needle biopsy should be performed to confirm the underlying histology, and radiographic surveillance is indicated until the time of surgery. The differential diagnosis of a renal mass includes cysts, benign neoplasms (adenoma, angiomyolipoma, oncocytoma), inflammatory lesions (pyelonephritis or abscesses), and other malignancies originating in the kidney such as transitional cell carcinoma of the renal pelvis, sarcoma, lymphoma, and Wilms' tumor or metastases from cancers originating in other organs. All of these are less common causes of renal masses than is RCC. The most common sites of distant metastases are the lungs, lymph nodes, liver, bone, and brain. These tumors may follow an unpredictable and protracted clinical course.

STAGING AND PROGNOSIS

Staging is based on the American Joint Committee on Cancer (AJCC) staging system (Fig. 85-1). Stage I tumors are ≤7 cm in greatest diameter and confined to the kidney, stage II tumors are >7 cm and confined to the kidney, stage III tumors extend through the renal capsule but are confined to Gerota's fascia, grossly infiltrate the renal vein, or involve regional lymph nodes (N1), and stage IV disease includes tumors that have invaded adjacent organs or involve nonregional lymph nodes or distant metastases. Sixty-five percent of patients present with stage I or II disease, 15–20% with stage III, and 15–20% with stage IV. The 5-year survival rate is currently 75% across all RCCs, but varies greatly by stage.

TABLE 85-2 Classification of Malignant Epithelial Neoplasms Arising from the Kidney

CARCINOMA TYPE	CHARACTERISTIC GROWTH PATTERN	CHROMOSOMAL EVENTS	GENES WITH RECURRENT SOMATIC ALTERATIONS
Clear cell	Acinar or sarcomatoid	3p-, 5q+, 14q-, 9p-	VHL, PBRM1, BAP1, SETD2
Papillary	Papillary or sarcomatoid	+7, +17, 9p-	MET, FH, CDKN2A (focal deletions)
Chromophobe	Solid, tubular, or sarcomatoid	Whole arm losses (1, 2, 6, 10, 13, 17, and 21)	TP53, PTEN, TERT promotor
Renal medullary carcinoma	Varying growth patterns, including cribriform, reticular, sarcomatoid, adenoid, and microcystic	+8q, 22q-, 22q translocations	SMARCB1 (focal deletions, mutations, gene fusions), SETD2
MITF translocation ^a	Mimicking clear cell and papillary variants	Xp11.2 translocations; t(6;11) translocations	TFE3 gene fusions, TFEB gene fusions

^aMicrophtalmia transcription factor gene family.

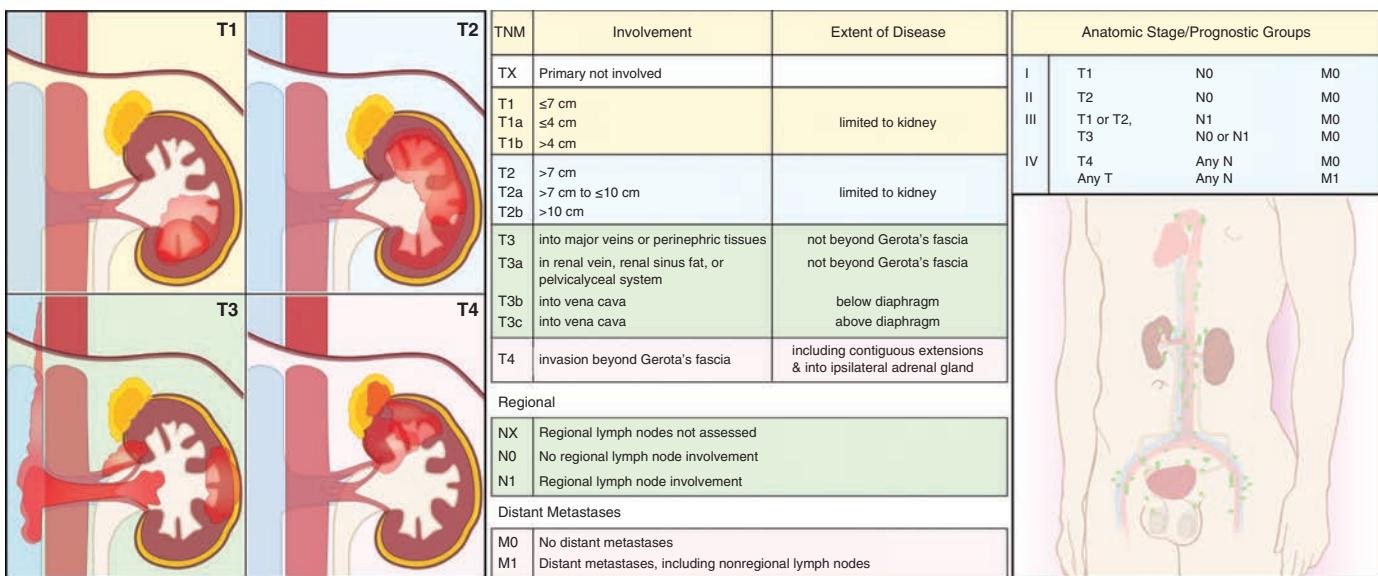


FIGURE 85-1 Renal cell carcinoma staging. TNM, tumor-node-metastasis.

Prognostic risk models are helpful for counseling patients diagnosed with metastatic disease and for anticipating survival rates when designing a clinical trial. A widely used prognostic model for advanced disease, the International Metastatic RCC Database Consortium (IMDC) risk model, incorporates six factors shown to correlate with worse survival: poor performance status, low hemoglobin concentration, high serum calcium, high neutrophil levels, high platelet levels, and <1 -year interval from diagnosis to systemic treatment. Patients with zero risk factors had significantly longer median survival (43 months) than patients with one or two risk factors (22.5 months) and those with three to six risk factors (8 months) when treated with first-line angiogenesis inhibitors (see below).

TREATMENT

Renal Cell Carcinoma

LOCALIZED TUMOR

The standard management for stage I or II tumors and selected cases of stage III disease is radical or partial nephrectomy. A radical nephrectomy involves en bloc removal of Gerota's fascia and its contents, including the kidney, and commonly the ipsilateral adrenal gland and regional lymph nodes that appear abnormal on imaging or intraoperatively. Open, laparoscopic, or robotic surgical techniques may be used. The role of a template lymphadenectomy in patients without apparent lymphadenopathy is controversial. Extension into the renal vein or inferior vena cava (stage III disease) does not preclude resection, which would then include thrombectomy.

Nephron-sparing approaches, i.e., open or laparoscopic partial nephrectomy, may be appropriate depending on the size and location of the tumor. This approach is particularly relevant for patients with solitary kidneys, bilateral tumors, or chronic renal insufficiency but can also be applied electively to resect small masses for patients with normal kidney function. Radical nephrectomy carries a greater risk for chronic kidney disease and cardiovascular morbidity and mortality.

Adjuvant systemic therapies, including cytokines and targeted agents, have been studied in randomized clinical trials, largely with negative results, and the standard of care remains active surveillance after nephrectomy.

METASTATIC DISEASE

Surgery has a limited role for patients with metastatic disease. Long-term survival may occur in patients who relapse with a solitary site

that is removed (metastasectomy). Nephrectomy despite presence of metastases (cytoreductive nephrectomy) is considered for carefully selected patients with stage IV disease. One indication for this approach can be to alleviate pain or hemorrhage of a primary tumor.

Radiation therapy is used for palliation of bone or brain metastases. The type of radiotherapy most commonly used is external-beam therapy, including stereotactic radiosurgery and other forms of image-guided radiotherapy.

Systemic therapy is the mainstay of care for metastatic disease. The timing of initiating such treatment should be carefully considered; some patients are asymptomatic at diagnosis, and with indolent behavior, it may be best to document progression before initiating treatment.

Metastatic RCC is refractory to cytotoxic chemotherapy. Patients are treated with molecularly targeted agents, including targeted immunotherapies. Treatments are continued with noncurative intent while tolerated and until disease progression is evident on cross-sectional imaging. Outcomes for patients with metastatic disease improved when increased understanding of underlying biology led to the successful development of several tyrosine kinase inhibitors (TKIs) targeting proangiogenic signaling through the VEGF receptors as well as allosteric inhibitors of mammalian target of rapamycin (mTOR) signaling. Serial large-scale randomized trials demonstrated that such agents, typically orally available, could be administered sequentially and in combination. Pivotal studies, by design, defined a dedicated space for each regimen in treatment-naïve or pretreated patients (**Table 85-3**).

Targeted immunotherapies were introduced after VEGF- and mTOR-directed agents had established standards of care in the first- and second-line setting. Nivolumab, a checkpoint inhibitor targeting PD-1, was compared to the mTOR inhibitor everolimus in a randomized trial in patients who had progressed on prior TKI therapy, challenging the standard approach in pretreated patients. Nivolumab demonstrated superior overall survival, positioning it as the new second-line agent of choice. Subsequently, immunotherapy combination regimens demonstrated efficacy in randomized trials conducted in treatment-naïve patients. In separate studies, two doublets demonstrated survival benefit over standard sunitinib therapy and changed the standard of care for untreated metastatic clear cell RCC: nivolumab in combination with the CTLA-4-directed checkpoint inhibitor ipilimumab proved superior to sunitinib in patients with high-risk features per the IMDC model, achieving complete radiographic disappearance of cancer in $>10\%$ of patients treated with the combination. In a second trial, the combination of the

TABLE 85-3 Commonly Used Systemic Regimens for Metastatic Renal Cell Carcinoma

CLASS	DRUG	FIRST FDA APPROVAL FOR RCC	CURRENTLY USED FOR
Antiangiogenic: TKIs	Sunitinib	2006	Advanced RCC, first line
	Pazopanib	2009	Advanced RCC, first line
	Axitinib	2012	Advanced RCC, pretreated
	Cabozantinib	2016 2017	Advanced RCC, pretreated with antiangiogenic therapy Advanced RCC, first line
Immunotherapy: checkpoint inhibitor	Nivolumab	2015	Advanced RCC, pretreated with antiangiogenic therapy
Combination therapies			
TKI + mTOR inhibitor	Lenvatinib + everolimus	2016	Advanced RCC, pretreated with one antiangiogenic therapy
PD-1 inhibitor + CTLA-4 inhibitor	Nivolumab + ipilimumab	2018	Advanced intermediate- or poor-risk RCC, first line
PD-1 inhibitor + TKI	Pembrolizumab + axitinib	2019	Advanced RCC, first line

Abbreviations: CTLA-4, cytotoxic T-lymphocyte-associated protein; FDA, U.S. Food and Drug Administration; mTOR, mammalian target of rapamycin; PD-1, programmed cell death-1; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor.

TKI axitinib together with the PD-1 inhibitor pembrolizumab was superior to sunitinib alone in all-comers with untreated metastatic RCC, again with high response rates across all IMDC risk groups. In both trials, responses were long-lasting, with improved time to disease progression and longer overall survival for combination regimens. Additional trials are ongoing to fortify the new standard of combination therapy in the first line.

With an ever-growing number of approved options directed toward different molecular targets, biomarkers are urgently needed to help individualize therapeutic choices and to gain insight as to whether and why treatments are working. Although a multitude of candidate biomarkers have been investigated for their predictive value in metastatic RCC, none have been validated for clinical use to date.

Projected overall survival in patients starting systemic therapies for newly diagnosed metastatic disease has tripled over the past 15–20 years; this can largely be attributed to the successful drug developments discussed here.

■ GLOBAL CONSIDERATIONS

Worldwide, over 400,000 patients are diagnosed each year with malignant tumors arising from the kidney, resulting in >175,000 deaths annually. Kidney cancer is the 10th most common cancer in men and the 15th most common cancer in women. Higher incidence is observed in developed countries, including the United States, Canada, Europe, Australia, New Zealand, and Uruguay. Relatively low rates are reported in Southeast Asia and Africa. The incidence of kidney cancer has been steadily increasing over the past four decades. Mortality trends have stabilized in Europe and the United States, but not in less developed countries. This is likely related to differences in access to optimal therapies. Treatment guidelines for both localized and metastatic renal cancer are similar between U.S. and European documents and contingent on the access to adequate health care and availability of targeted drugs to treat metastases.

■ FURTHER READING

- CHOUERI TK, MOTZER RJ: Systemic therapy for metastatic renal cell carcinoma. *N Engl J Med* 376:354, 2017.
- MOCH H et al: The 2016 WHO classification of tumours of the urinary system and male genital organs—Part A: renal, penile, and testicular tumours. *Eur Urol* 70:93, 2016.
- MOTZER RJ et al: Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med* 378:1277, 2018.
- MOTZER RJ et al: Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade. *Cancer Cell* 38:803, 2020.
- RINI BI et al: Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med* 380:1116, 2019.
- SANCHEZ A et al: Current management of small renal masses, including patient selection, renal tumor biopsy, active surveillance, and thermal ablation. *J Clin Oncol* 36:3591, 2018.

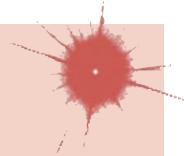
TURAJLIC S et al: Tracking cancer evolution reveals constrained routes to metastases: TRACERx Renal. *Cell* 173:581, 2018.

WONG MCS et al: Incidence and mortality of kidney cancer: Temporal patterns and global trends in 39 countries. *Sci Rep* 7:15698, 2017.

86

Cancer of the Bladder and Urinary Tract

Noah M. Hahn



GLOBAL CONSIDERATIONS

Within the United States, urothelial carcinomas of the bladder and urinary tract are most closely related to tobacco smoking history. However, within developing countries, water supplies contaminated with arsenic or schistosomiasis parasites also are major carcinogenic contributors.

INTRODUCTION

Cancers of the urinary tract including the bladder, renal pelvis, ureter, and urethra occur frequently, and they represent the second most common class of genitourinary cancers. Bladder cancer alone represents the sixth most common cancer diagnosis annually in the United States with 81,400 new cases and 17,980 deaths every year. Because cancers of the renal pelvis are often lumped in with all kidney cancers, the true incidence and mortality from nonbladder urinary tract cancers are less precise. While less frequent than bladder cancer, an additional 20,000 new cases and 5000 deaths are estimated every year. An accelerated understanding of the molecular underpinnings of bladder and urinary tract cancer biology has led to a significant increase in urothelial cancer clinical trials resulting in U.S. Food and Drug Administration (FDA) approval of multiple new therapeutic agents since 2016 with more expected to follow. This chapter reviews the established, current, and emerging evidence that serves as the basis for the rapidly evolving standards of care for patients with bladder and urinary tract cancers.

■ CLINICAL EPIDEMIOLOGY AND RISK FACTORS

Bladder cancer typically affects older patients with a median age at diagnosis of 73 years. Males are four times more frequently affected than females. Similarly, bladder cancer is more common in Caucasians than in Asian patients. Inheritable germline genetic risk factors have been identified in up to one-seventh of patients with bladder or urinary tract cancers. However, a singular germline genetic alteration has not

been observed in a majority of these cases, and the impact of germline genetic alterations on family members of urothelial cancer patients is uncertain. Patients with defects in mismatch repair genes leading to microsatellite instability (*MLH1*, *MSH2*, *MSH6*, etc.) as part of the familial cancer Lynch syndrome are at particular risk of upper urinary tract cancers of the renal pelvis and ureter. Additionally, patients with Cowden disease (*PTEN* mutations) or retinoblastoma (*RB1* mutations) are at increased risk for developing bladder cancer.

Historically, associations have existed between environmental toxic exposures and higher rates of developing bladder cancer. Carcinogenic agents associated with increased risk of bladder cancer have included the aromatic amines benzidine and beta-naphthylamine that can be present in industrial dyes as well as arsenic that can be found in some drinking water supplies in underdeveloped countries. Other chemicals in the leather, paint, rubber, textiles, and printing industries have been associated with bladder cancer. Associations with exposures to hair dyes and hair sprays in workers in the hairstyling field have been suggested. Additionally, concern has been raised regarding use of the antidiabetic medication, pioglitazone, and bladder cancer risk. Extensive reviews and meta-analyses have produced differing conclusions. The data suggest a small risk of bladder cancer from long-term pioglitazone use, which has led to inclusion of bladder cancer risk within the pioglitazone prescribing information. An association between chronic inflammatory states and the development of squamous bladder cancer clearly exists in underdeveloped countries in patients chronically infected with the parasitic disease schistosomiasis and in paraplegic patients with chronic indwelling catheters. Above and beyond each of these associations, however, smoking of tobacco products (cigarettes, cigars, pipes, etc.) remains the overwhelming leading risk factor for development of bladder cancer. Among new bladder cancer diagnoses, 90% of cases occur in current or former smokers. Toxicologists have estimated that >70 confirmed carcinogenic toxins are present within tobacco smoke. It is estimated that one-third of bladder cancer cases could be prevented through simple modification of lifestyle choices, in particular cessation of smoking.

■ CLINICAL PRESENTATION AND DIAGNOSTIC WORKUP

Occasionally, patients will present with flank pain in association with an upper tract renal pelvis or ureter cancer or due to hydronephrosis in association with a bladder tumor obstructing the orifice of the ureter within the bladder. Only in rare cases do patients present with significant cachexia and widespread metastatic disease. For most patients, painless hematuria (either gross or microscopic) represents the initial manifestation of an underlying urinary tract cancer. In females, hematuria due to malignancy can often be mistaken for a urinary tract infection or menstrual bleeding. While treatment with antibiotics is warranted if a concurrent urinary tract infection is noted on initial urinalysis, persistent hematuria requires further workup. Painless hematuria in males is almost always abnormal and should be worked up. Initial investigations in patients of either sex should include urine cytology and visual examination of the bladder by cystoscopy. Cytology is successful in identifying cancer in only 50% of individuals with high-grade bladder cancers. In addition to urine cytology, radiographic evaluation of the kidneys and upper urinary tract by CT urogram should be performed. Because of the increased sensitivity and reduced IV contrast loads, CT uograms have largely replaced IV pyelograms as the preferred upper urinary tract imaging modality. A magnetic resonance (MR) urogram may be substituted in patients with poor renal function. Additional diagnostic testing of the urine to assess for cancer-associated chromosomal changes by fluorescent *in situ* hybridization, increased levels of nuclear mitotic proteins, increased bladder tumor-associated antigens, or higher levels of staining on cells shed by the bladder may identify some cancers missed by traditional cytology testing. However, they may also produce abnormal results in patients who do not have cancer. For now, these adjunct molecular tests are primarily utilized in detecting recurrent cancer in patients with a prior diagnosis of urinary tract cancer. Small tumors, particularly flat noninvasive tumors of the bladder, may be detected at higher rates with

the use of blue light cystoscopy or narrow-band imaging cystoscopy. Both blue light and narrow-band imaging cystoscopies are now used routinely in the monitoring of patients with bladder cancer. For patients with no bladder abnormalities in whom upper tract tumors are suspected, visualization of the upper urinary tracts and renal pelvices should be performed by ureteroscopy or retrograde pyelography.

In all patients with abnormalities noted in the bladder or upper urinary tracts, complete endoscopic resection for histologic diagnosis and staging should be performed when possible via either transurethral resection of bladder tumor (TURBT) or endoscopic resection of upper tract tumors.

■ HISTOLOGY

Urothelial carcinoma, often called transitional cell carcinoma in the past, is the most common urinary tract cancer histology and is observed in ~90% of cases. Squamous, glandular, micropapillary, plasmacytoid, sarcomatoid, and other variant features can often be found in portions of urothelial carcinoma tumors; however, pure variant histologies are rare. The presence of some variant histologies including micropapillary and plasmacytoid has been associated with worse surgical outcomes compared to urothelial carcinoma. Nonurothelial variant histologies including squamous cell carcinoma, adenocarcinoma, small-cell carcinoma, and carcinosarcoma collectively account for ≤10% of urinary tract tumors. Examples of traditional urothelial carcinoma and some of the variant histologies are shown in Fig. 86-1.

■ MOLECULAR BIOLOGY

Clinically, urothelial carcinoma of the bladder displays a biphasic phenotype characterized by (1) low-grade papillary tumors that frequently recur but rarely invade or metastasize and (2) high-grade sometimes flat tumors that invade early leading to lethal metastatic disease. In both of these phenotypes, loss of portions of chromosomes 9q and 9p by loss of heterozygosity is an early molecular event, whose exact significance is not clear. Potential candidate regulatory genes in these genomic regions include *CDNK2A*, a cyclin-dependent kinase inhibitor, and *TSC1*, a gene encoding hamartin mutated in tuberous sclerosis. Early investigations have demonstrated that low-grade tumors are characterized by alterations in the *RAS/RAF* signaling pathway with activating *FGFR3* mutations or gene fusions present in 60–80% of patients. In contrast, the high-grade invasive phenotype is notable for early deleterious mutations in *TP53* and *RB1*, alterations in *CDH1* (E-cadherin), and increased expression of *VEGFR2*. In urothelial carcinoma of the renal pelvis and ureter, 10–20% of cases may be associated with Lynch syndrome hereditary defects in the *MLH1*, *MSH2*, or *MSH6* mismatch repair genes, leading to microsatellite instability and frequent DNA mutations. Testing for germline mutations in these genes is recommended in patients with upper urinary tract urothelial carcinoma under the age of 60 at diagnosis, with a first-degree relative with a Lynch syndrome-associated cancer diagnosed under the age of 50, or with two first-degree relatives with a Lynch syndrome-associated cancer regardless of the age at diagnosis.

As genomic analysis technologies have improved, so has our understanding of the molecular biology unique to urothelial carcinoma. In 2017, the full bladder cancer results of The Cancer Genome Atlas (TCGA) project were published. This effort comprehensively analyzed gene mutations, fusions, expression, copy number variations, methylation, and microRNA across the genome of patients with bladder urothelial carcinoma treated with surgery. Key findings from this effort include (1) genomic alterations in genes (e.g., *FGFR3*, *EGFR*, *ERBB2*, *ERBB3*, *PIK3CA*, *TSC1*, etc.) targetable by currently approved drugs or drugs in development in 71% of patients; (2) genomic alterations in chromatin-modifying genes (*KMT2D*, *KDM6A*, *KMT2C*, *EP300*, *CREBBP*, etc.) in the majority of patients; (3) hypermethylation with epigenetic silencing of gene expression in one-fourth of patients; and (4) the identification by RNA sequencing of five distinct intrinsic molecular subtypes (luminal papillary, luminal infiltrated, luminal, basal squamous, and neuronal) closely resembling luminal and basal subclassifications of breast cancers. These bladder TCGA findings have led to clinical trial designs enriching for patients with specific

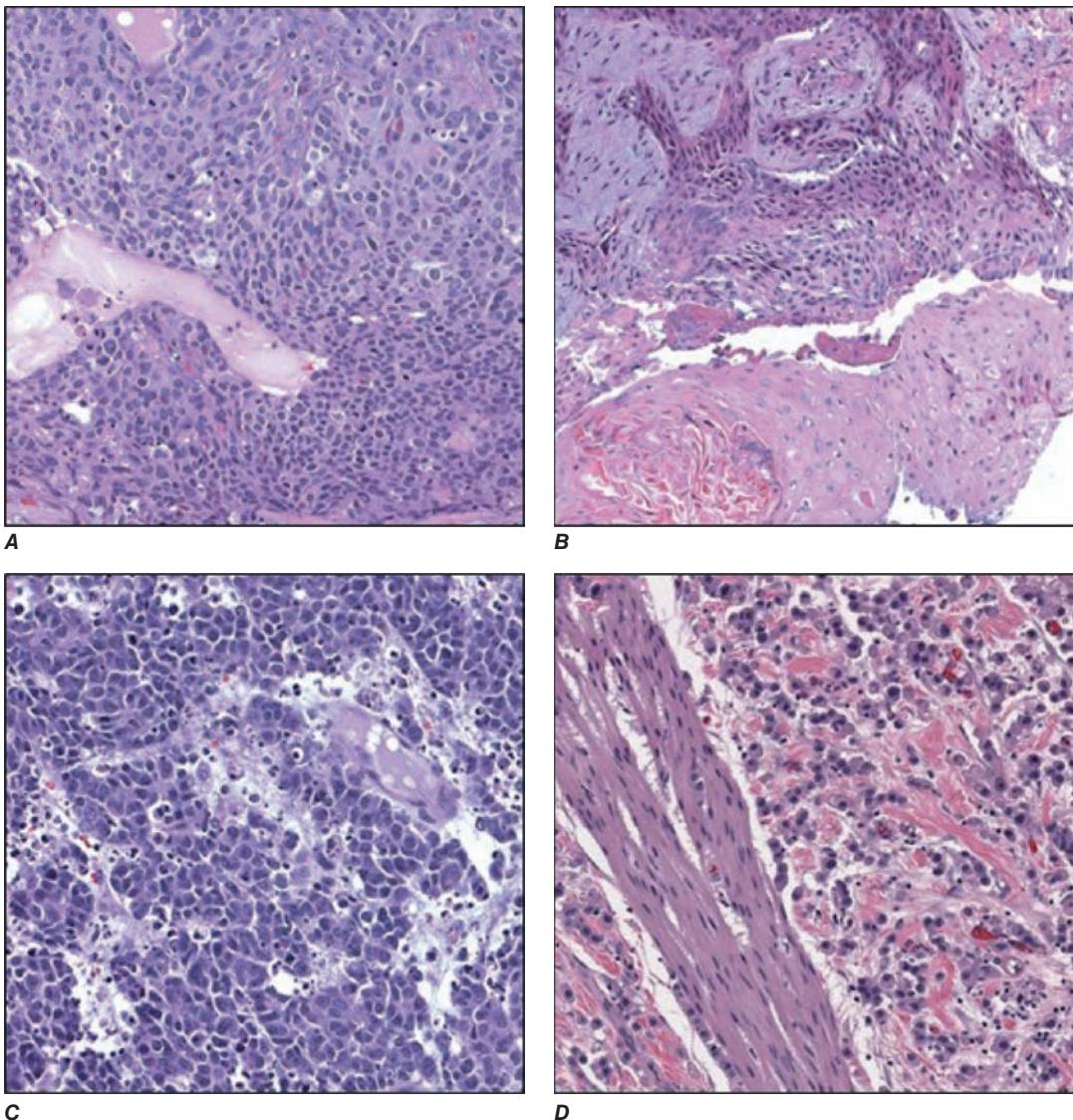


FIGURE 86-1 Bladder and urinary tract cancer histologies. **A.** Urothelial carcinoma. **B.** Squamous cell carcinoma. **C.** Small-cell carcinoma. **D.** Plasmacytoid variant. (Courtesy of Alex Baras, MD, PhD, Johns Hopkins University Department of Pathology.)

gene mutation profiles as well as interrogation of candidate biomarkers according to intrinsic molecular subtypes.

■ STAGING AND OUTCOMES BY STAGE

The staging of bladder cancer is dependent on the depth of invasion within the bladder wall, involvement of lymph nodes, and spread to surrounding and distant organs as depicted in Fig. 86-2. Approximately 75% of bladder cancer presents with non-muscle-invasive bladder cancer (NMIBC), 18% with disease invading into or through the muscular wall of the bladder, and only 3% presenting with metastatic spread to distant organs. NMIBC is defined by tumors that involve only the immediate epithelial layer of cells (carcinoma in situ [CIS] and Ta) or that only penetrate into the connective tissue below the urothelium (T1) but not into the muscular layer known as the *muscularis propria*. Muscle-invasive bladder cancer (MIBC) is defined by tumors that invade into the muscularis propria (T2), through the muscularis propria to involve the surrounding serosa (T3), or into immediately adjacent pelvic organs such as the rectum, prostate, vagina, or cervix (T4). Lymph node staging is classified according to involvement of a solitary node within the true pelvis (N1), two nodes involved in the true pelvis (N2), or involvement of the common iliac nodes (N3). Any disease that has spread beyond the common iliac nodes is considered metastatic (M1). The staging of bladder cancer is driven primarily by the T stage of the tumor, with stages 0a-II defined entirely by the T stage in the absence of nodal or metastatic disease. Involvement of

regional lymph nodes in the true pelvis or along the common iliac artery qualifies as stage III disease, whereas involvement of any distant metastases qualifies as stage IV disease. Clinical outcomes of patients with bladder cancer correlate closely with staging at diagnosis with 5-year overall survival rates of 70–90% for disease confined to the bladder (stage I-II), 36–50% for disease that penetrates through the bladder or has spread to regional lymph nodes (stage III), and only 5% for disease extending to metastatic sites (stage IV).

■ TREATMENT APPROACHES

Early-Stage Disease For NMIBC, removal of all visible tumors by TURBT in the operating room is considered the mainstay of surgical treatment. Risk of recurrence can be classified as low, intermediate, or high depending on the presence of features summarized in Table 86-1. For patients with low-risk disease, meta-analyses have demonstrated a 12% reduction in early relapses when a single chemotherapy treatment of mitomycin C, epirubicin, or gemcitabine was instilled directly into the bladder (intravesical therapy) within 24 hours of the TURBT. For patients with intermediate- or high-risk tumors, weekly intravesical instillations for 6 consecutive weeks of the attenuated mycobacterium strain known as *Bacille Calmette-Guérin* (BCG) reduce the risk of recurrence at 12 months from 56 to 29%. In addition, BCG treatment has been shown to decrease the rate of progression to MIBC by 27%. Intravesical

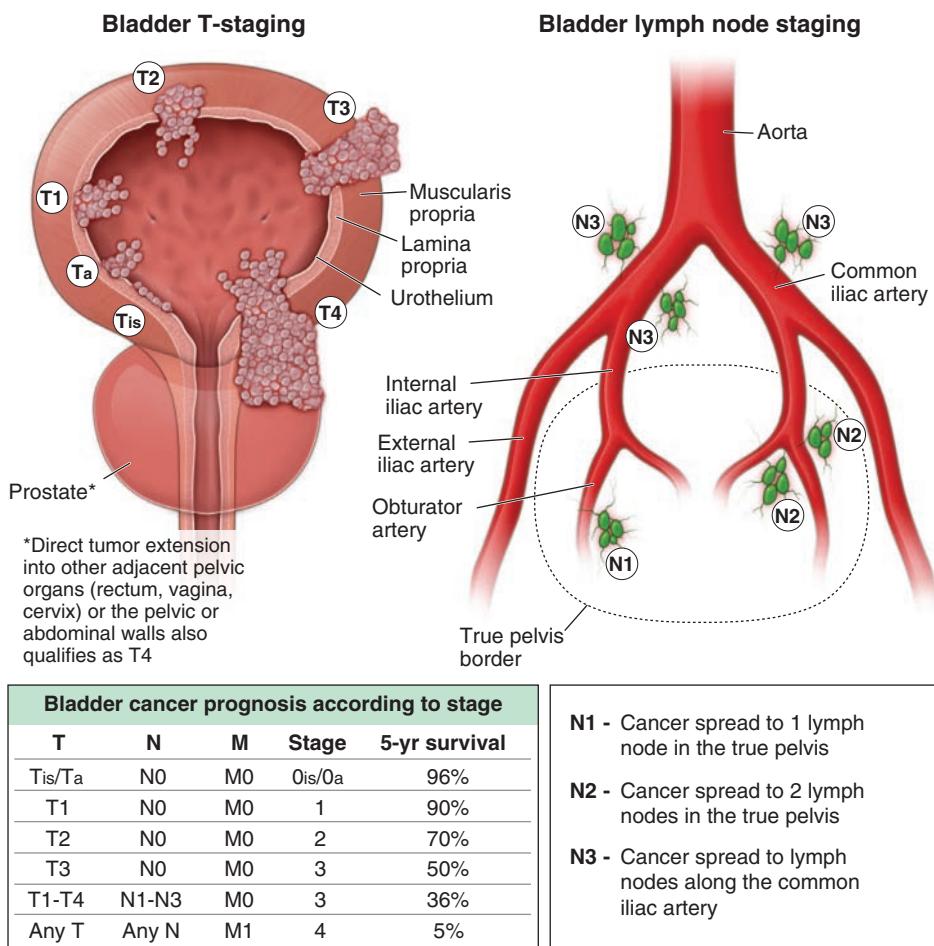


FIGURE 86-2 Bladder cancer staging and prognosis. TNM, tumor-node-metastasis.

BCG is generally well tolerated. Side effects can include dysuria, urinary frequency, bladder spasms, hematuria, and, in rare cases (<5%), a systemic inflammatory response that can mimic disseminated BCG infection. Following a 6-week induction BCG schedule, additional maintenance BCG treatments given according to the Southwest Oncology Group schedule further reduce the risk of recurrent NMIBC compared to induction BCG alone. In patients with NMIBC that recurs long after initial BCG treatment, a repeat course of BCG can be considered. For patients with recurrence after a second adequate course of BCG or with relapsed NMIBC within 6 months of initial BCG exposure, surgical removal of the entire bladder by cystectomy is recommended due to the high risk of progression to MIBC and potentially metastatic disease. For patients who are not fit enough for or who refuse cystectomy, non-BCG alternative intravesical agents (mitomycin C, gemcitabine, docetaxel, valrubicin) or systemically

administered agents that inhibit the PD-1/PD-L1 immune checkpoint pathway (pembrolizumab) can achieve durable tumor responses in a small fraction of patients.

Upper Tract Disease In patients with urothelial carcinoma of the renal pelvis or ureter, endoscopic tissue acquisition and staging are more challenging than primary tumors located in the bladder. Tumors possessing all of the following are considered low risk: solitary tumor, low grade, size <1 cm, and no invasive component on imaging. Low-risk tumors can successfully be treated by laser ureteroscopic ablation or surgical resection and reanastomosis of the remaining ureter ends in tumors that cannot be successfully eradicated endoscopically.

Muscle-Invasive Disease In patients with urothelial carcinoma of the bladder that invades into or through the muscularis propria but with no evidence of metastatic spread, more aggressive therapy options summarized in **Table 86-2** are required to achieve cure. In carefully selected patients with no evidence of CIS or hydronephrosis, bladder-sparing combined-modality therapy with concurrent chemotherapy and radiation can achieve cure in ~65% of patients. Various chemotherapy regimens have been utilized in combination with radiation including cisplatin, carboplatin, 5-fluorouracil, mitomycin C, paclitaxel, and gemcitabine. It is important to note that a maximal debulking of all visible

tumor by TURBT is required prior to initiation of combined-modality therapy. In patients who achieve a complete response to combined-modality therapy, regular cystoscopic monitoring of the bladder is required with salvage cystectomy offered to patients who develop MIBC in follow-up.

In a similar fashion, bladder-sparing partial cystectomy can be performed in a very small subset of MIBC patients. The ideal patient for partial cystectomy is the patient with a solitary, clinical T2 urothelial carcinoma in the dome of the bladder. In such patients, the tumor and immediate surrounding urothelium can be resected with reconstruction of the remaining bladder to maintain near physiologic urinary function.

In the majority of patients, however, resection of the entire bladder is required. In males, a cystoprostatectomy with removal of the bladder, prostate, and pelvic lymph nodes is performed, whereas in females, an anterior exenteration with removal of the bladder, uterus, ovaries, cervix, and pelvic lymph nodes is performed. With the bladder removed, three options exist to reroute the urine outflow. In an ileostomy, the bilateral ureters are connected to a portion of ileum that is brought through an incision in the abdominal wall to create a stoma that drains urine into an affixed bag outside of the body. In a continent urinary reservoir or “Indiana pouch,” the ureters are connected to a portion of ileum that has been separated on both ends from the rest of the small-bowel transit to form a urinary reservoir. The remaining small bowel is reanastomosed, and the urinary reservoir is brought up just beneath the abdominal wall muscles with patients catheterizing the urinary reservoir several times per day via a small stoma tract. Last, in a neobladder, the same urinary reservoir described previously is brought down into the pelvis and is anastomosed to the remaining urethra to provide the opportunity to the patient to void urine through the urethra. The choice of which urinary reconstruction to perform is affected not only by patient choice but also by anatomic tumor considerations and

TABLE 86-1 Non-Muscle-Invasive Bladder Cancer Recurrence Risk Groups

RISK GROUP	CHARACTERISTICS
Low risk	Initial tumor, solitary tumor, low grade, <3 cm, no CIS
Intermediate risk	All tumors not defined in the two adjacent categories (between the category of low and high risk)
High risk	Any of the following: <ul style="list-style-type: none"> T₁ tumor High-grade CIS Multiple and recurrent and large (>3 cm) Ta low-grade tumors (all conditions must be met for this point on Ta low-grade tumors)

Abbreviation: CIS, carcinoma in situ.

TABLE 86-2 Treatment Approaches to MIBC Patients

TREATMENT	PATIENT SELECTION	CLINICAL OUTCOMES
Bladder-sparing chemoradiation	No CIS, no hydronephrosis, maximal TURBT required	65% cure, 55% bladder intact, highly dependent on patient selection
Bladder-sparing partial cystectomy	Solitary tumors in dome of bladder are ideal	Variable, highly dependent on patient selection
Cystectomy	Any MIBC patient	50% cure with surgery alone, highly dependent on pathologic stage
Neoadjuvant cisplatin-based chemotherapy	Cisplatin-eligible MIBC patients	5–10% improvement in overall survival compared to cystectomy alone
Adjuvant cisplatin-based chemotherapy	Cisplatin-eligible high-risk postcystectomy MIBC patients (pT3-4, N+)	Similar improvement as neoadjuvant treatment, data less robust, many patients not suitable for adjuvant treatment

Abbreviations: CIS, carcinoma in situ; MIBC, muscle-invasive bladder cancer; TURBT, transurethral resection of bladder tumor.

urologist experience with each procedure. Regardless of the type of surgery performed, all patients undergo a significant catabolic change in their metabolism following removal of the bladder. While many MIBC patients are affected by weight loss preoperatively, it is not uncommon for postcystectomy patients to lose an additional 10–15 lb in the first month postoperatively. In addition, patients can experience long-term nutritional changes such as low B_{12} levels due to alterations in small-bowel physiology caused by all of the urinary diversion options.

Despite aggressive surgery, only half of patients undergoing cystectomy are cured by surgery alone. Therefore, many clinical trials have investigated the role of systemic chemotherapy before (neoadjuvant) or after (adjuvant) surgery. Meta-analyses have shown a 5–10% absolute overall survival advantage when combination chemotherapy regimens utilizing cisplatin have been used before surgery. A similar benefit exists with cisplatin-based combination chemotherapy given after surgery. However, the data in the adjuvant setting are based on smaller, older trials. Furthermore, in the postoperative setting, some patients may not recover sufficiently from their surgery within a time frame optimal for chemotherapy administration. Importantly, non-cisplatin-containing chemotherapy regimens have proven inferior to cisplatin-containing regimens. Therefore, if patients are not suitable candidates for cisplatin administration due to poor functional status or comorbidities (e.g., poor renal function), patients should proceed directly to surgery and forego neoadjuvant therapy.

For patients with high-risk urothelial carcinoma of the upper urinary tract, resection of the kidney and ureter (including the ureter-bladder cuff) by nephroureterectomy is preferred. Segmental ureterectomy may be appropriate in patients with decreased renal function in which nephron-sparing outcomes are critical to prevent the need for dialysis. Similarly, in CIS patients, administration of BCG therapy via a nephrostomy tube can be considered to preserve intact renal function. The use of cisplatin-based neoadjuvant chemotherapy has been associated with a pathologic complete response at surgery of 14% in upper tract urothelial carcinoma patients. Similarly, in the post-nephroureterectomy setting, adjuvant platinum-based chemotherapy (carboplatin or cisplatin) reduced recurrence rates by 55% compared to surgery alone. The use of perioperative chemotherapy either before or after surgery is now recommended for upper tract urothelial carcinoma patients in national guidelines.

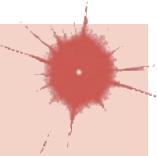
Metastatic Disease For patients with metastatic urothelial carcinoma regardless of primary tumor origin, systemic chemotherapy is the most established standard of care. In a randomized phase 3 clinical

trial, the combination of methotrexate, vinblastine, doxorubicin, and cisplatin (MVAC) demonstrated an improvement in median overall survival from 8.2 to 12.5 months compared to single-agent cisplatin. In a head-to-head randomized phase 3 clinical trial, the combination of cisplatin and gemcitabine (CG) demonstrated similar overall survival compared to MVAC with a more favorable side effect profile. Since 2000, treatment with either MVAC or CG has remained a standard first-line treatment of patients with metastatic urothelial carcinoma with adequate renal function and functional status suitable for cisplatin therapy. For patients with lymph node-only metastases and good functional status, cure is achieved in 15–20% of such patients. Unfortunately, only ~5% of metastatic patients fulfill both these criteria. Furthermore, approximately half of patients with urothelial carcinoma have renal insufficiency, comorbidities, or frail functional status, and are not candidates for cisplatin treatment. In cisplatin-ineligible patients, carboplatin-based chemotherapy regimens have historically been used with median overall survival rates decreased to 9.3 months. Agents inhibiting the immune checkpoint programmed cell death protein 1 (PD-1) and programmed death ligand 1 (PD-L1) pathways have become additional standard options for both front-line chemotherapy-naïve (atezolizumab, pembrolizumab), front-line maintenance (avelumab), and second-line postplatinum (pembrolizumab, nivolumab, and avelumab) metastatic urothelial carcinoma patients. Although these agents only result in tumor responses in 10–30% of patients, they have been approved due to their improved safety profiles compared to traditional chemotherapy options and the prolonged durability of some tumor responses. These agents aim to reactivate a patient's own immune system to recognize and eliminate their cancer. As such, their unique side effect profile is characterized by immune-related toxicities that are rare but can be severe and may include colitis, pneumonitis, hepatitis, nephritis, myocarditis, rash, hypothyroidism, Guillain-Barré syndrome, idiopathic thrombocytopenia purpura, and adrenal insufficiency.

In patients with activating tumor fibroblast growth factor 2 or 3 (FGFR2/3) mutations or fusions with progressive disease following platinum-based therapy, the oral FGFR tyrosine kinase inhibitor erdafitinib is another standard option resulting in tumor responses in 32% of patients with a median duration of response of 5.4 months. Additionally, thenectin-4-targeting antibody-drug conjugate enfortumab vedotin provides an additional standard option for patients with progression after both platinum-based therapy and PD-1/PD-L1 immune checkpoint therapy independent of tumor mutation status. Tumor responses are observed in 44% of patients, including patients with liver metastases, with a median response duration of 7.6 months. Additional novel urothelial carcinoma therapeutics are under ongoing investigation.

FURTHER READING

- AMERICAN CANCER SOCIETY: *Cancer Facts & Figures 2020*. Atlanta, GA: Available from <https://www.cancer.org/cancer/bladder-cancer/detection-diagnosis-staging/survival-rates.html>.
- CARLO MI et al: Cancer susceptibility mutations in patients with urothelial malignancies. *J Clin Oncol* 38:5, 2020.
- HOWLADER N et al: SEER Cancer Statistics Review, 1975–2017. Available from https://seer.cancer.gov/csr/1975_2017/. Based on November 2019 SEER data submission, posted to the SEER website, April 2020.
- KAMAT AM et al: Bladder cancer. *Lancet* 388:2796, 2016.
- KNOWLES MA et al: Molecular biology of bladder cancer: New insights into pathogenesis and clinical diversity. *Nat Rev Cancer* 15:25, 2015.
- ROBERTSON AG et al: Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* 171:3, 2017.
- SANTOPIETRO AL et al: Advances in the management of urothelial carcinoma: Is immunotherapy the answer? *Expert Opin Pharmacother* 22:1743, 2021.
- SIEGEL RL et al: Cancer statistics, 2020. *CA Cancer J Clin* 70:1, 2020.



Benign and malignant changes in the prostate increase with age. Autopsies of men in the eighth decade of life show hyperplastic changes in >90% and malignant changes in >70% of individuals. The high prevalence of these diseases among the elderly, who often have competing causes of morbidity and mortality, mandates a risk-adapted approach to diagnosis and treatment. This can be achieved by considering these diseases as a series of states. Each state represents a distinct clinical milestone for which therapy(ies) may be recommended based on disease extent, current symptoms, the risk of developing symptoms, or the risk of death from disease in relation to death from other causes within a given time frame. For benign proliferative disorders, symptoms of bladder outlet obstruction and potential complications including urinary retention and urinary tract infection are weighed against the side effects and complications of medical or surgical intervention. For prostate malignancies, the likelihood that a clinically significant cancer is present in the gland and the concomitant risk of symptoms or death from cancer are balanced against the morbidities of the recommended treatments and preexisting comorbidities.

ANATOMY AND PATHOLOGY

The prostate is in the pelvis and is adjacent to the rectum, the bladder, the periprostatic and dorsal vein complexes, the neurovascular bundles that are responsible for erectile function, and the urinary sphincter that is responsible for passive urinary control. The prostate is composed of branching tubuloalveolar glands arranged in lobules surrounded by fibromuscular stroma. The acinar unit includes an epithelial compartment made up of epithelial, basal, and neuroendocrine cells separated by a basement membrane and a stromal compartment that includes fibroblasts and smooth-muscle cells. Prostate-specific antigen (PSA) and prostatic acid phosphatase (PAP) are produced in the epithelial cells. Both prostate epithelial cells and stromal cells express androgen receptors (ARs) and depend on androgens for growth. Testosterone, the major circulating androgen, is converted by the enzyme 5 α -reductase to dihydrotestosterone in the gland.

The periurethral portion of the gland increases in size during puberty and after the age of 55 years due to the growth of nonmalignant cells in the transition zone of the prostate that surrounds the urethra. Most cancers develop in the peripheral zone, and cancers in this location may be palpated during a digital rectal examination (DRE).

PROSTATE CANCER

The American Cancer Society's estimates for prostate cancer in the United States for 2021 are ~248,530 new prostate cancer cases and ~34,130 deaths from prostate cancer. The absolute number of prostate cancer deaths has decreased in the past 10 years, attributed by some to the widespread use of PSA-based detection strategies. However, the paradox of management is that although 1 in 8 men will eventually be diagnosed with prostate cancer and the disease remains the second leading cause of cancer deaths in men, only 1 man in 41 with prostate cancer will die of his disease.

EPIDEMIOLOGY

Epidemiologic studies show that the risk of being diagnosed with prostate cancer increases 2.5-fold if one first-degree relative is affected and fivefold if two or more are affected. Current estimates are that 40% of early-onset and 5–10% of all prostate cancers are hereditary. Prostate cancer affects ethnic groups differently. Matched for age, African-American males have a higher incidence and present at a more advanced stage with higher-grade, more aggressive cancers. Genome-wide association studies (GWAS) have identified >40 prostate cancer susceptibility loci that are estimated to explain up to 25% of prostate

cancer risk. Among the genes implicated in variations in incidence and outcome are single-nucleotide polymorphisms (SNPs) in the vitamin D receptor in African Americans and variants in the AR, CYP3A4, both involved in the deactivation of testosterone, as well as CYP17, which is involved in steroid biosynthesis. One early change is hypermethylation of the GSTP1 gene promoter, which leads to loss of function of a gene that detoxifies carcinogens. The finding that many prostate cancers develop adjacent to a lesion termed *proliferative inflammatory atrophy* (PIA) suggests a role for inflammation.

The prevalence of autopsy-detected cancers is similar around the world, while the incidence of clinical disease varies. Thus, environmental and dietary factors may play a role in prostate cancer growth and progression. High consumption of dietary fats, such as α -linoleic acid or polycyclic aromatic hydrocarbons that form when red meats are cooked, is believed to increase risk. Like breast cancer in Asian women, the risk of prostate cancer in Asian men increases when they move to Western environments. Protective factors include consumption of the isoflavonoid genistein (which inhibits 5 α -reductase), cruciferous vegetables with isothiocyanate sulforaphane, lycopene found in tomatoes, and inhibitors of cholesterol biosynthesis (e.g., statin drugs). Not smoking, regular exercise, and maintaining a healthy body weight may reduce the risk of progression.

DIAGNOSIS AND TREATMENT BY CLINICAL STATE

The prostate cancer continuum—from the appearance of a preneoplastic and invasive lesion that is localized to the gland, to a metastatic lesion causing symptoms and, ultimately, mortality—can span decades. To limit overdiagnosis of clinically insignificant cancers and for disease management in general, competing risks are considered in the context of a series of clinical states (Fig. 87-1). The states are defined operationally based on whether or not a cancer diagnosis has been established and, for those with a diagnosis, the state of the primary tumor (treated vs untreated), whether or not metastases are detectable on imaging studies, and the measured level of testosterone in the blood. With this approach, an individual resides in only one state and remains in that state until he has progressed. At each assessment, the decision to offer treatment and the specific form of treatment are based on the presence or absence of cancer-related symptoms, and if absent, the risk posed by the cancer relative to competing causes of morbidity and mortality that may be present in that individual. It follows that the more advanced the disease, the greater is the need for treatment.

For those without a cancer diagnosis, the decision to undergo testing to detect a cancer is based on the individual's estimated life expectancy and, separately, the probability that a clinically significant cancer may be present. For those with a prostate cancer diagnosis, the clinical states model considers the probability of developing symptoms or dying from the disease. Thus, a patient with a localized tumor that has been surgically removed remains in the state of localized disease if the PSA remains undetectable. The time within a state then becomes a measure of the efficacy of an intervention, though the effect may not be assessable for years. Because many men with active cancer are not at risk for developing metastases, symptoms, or death, the clinical states model allows a distinction between *cure*—the elimination of all cancer cells, the primary therapeutic objective of treatment for most cancers—and *cancer control*, by which the tempo of the illness is determined to be so slow or has been altered by treatment to the point where it is unlikely to cause symptoms, to metastasize, or to shorten a patient's life expectancy. Importantly, from a patient standpoint, both outcomes can be considered equivalent therapeutically assuming the patient has not experienced symptoms of the disease or the treatment needed to control it. Even when a recurrence is documented, immediate therapy is not always necessary. Rather, as at the time of diagnosis, the need for intervention is based on the tempo of the illness as it unfolds in the individual, relative to the risk-to-benefit ratio of the intervention being considered.

NO CANCER DIAGNOSIS

Prevention No agent is currently approved for the prevention of prostate cancer. The results from several large double-blind,

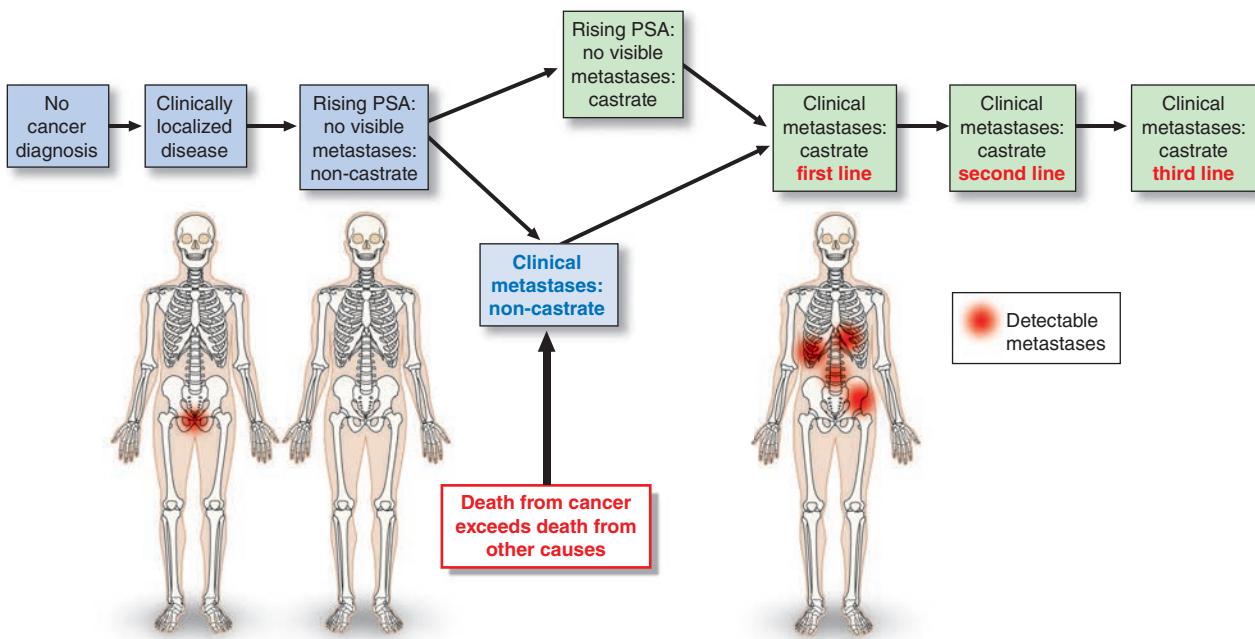


FIGURE 87-1 Clinical states of prostate cancer. PSA, prostate-specific antigen.

randomized chemoprevention trials have established 5 α -reductase inhibitors (5ARIs) as the predominant therapy to reduce the future risk of a prostate cancer diagnosis. The Prostate Cancer Prevention Trial (PCPT), in which men aged >55 years received placebo or the 5ARI finasteride, which inhibits the type 1 isoform, showed a 25% (95% confidence interval 19–31%) reduction in prostate cancer incidence from 24% with placebo to 18% with finasteride. In REDUCE (Reduction by Dutasteride of Prostate Cancer Events trial), a reduction in incidence from 25% with placebo to 20% with dutasteride was found ($p = .001$). Dutasteride inhibits both the type 1 and type 2 5ARI isoforms. While both studies met their endpoint, there was concern that most of the cancers that were “prevented” were low risk. Neither drug is approved for prostate cancer prevention. In comparison, the Selenium and Vitamin E Cancer Prevention Trial (SELECT), which enrolled African-American men aged ≥ 50 years and others aged ≥ 55 years, showed no difference in cancer incidence in patients receiving vitamin E (4.6%) or selenium (4.9%) alone or in combination (4.6%) relative to placebo (4.4%). A similar lack of benefit for vitamin E, vitamin C, and selenium was seen in the Physicians Health Study II.

Early Detection and Diagnosis The decision to pursue a diagnosis of prostate cancer must balance the benefit from detecting and treating clinically significant cancers that, left untreated, would adversely affect a patient's quality and duration of life, against the morbidity associated with the overdiagnosis and overtreatment of clinically insignificant cancers that are highly prevalent in the general population. The balance is best approached through shared decision-making between the patient and physician. Considerations for whether to pursue a diagnosis include symptoms, an abnormal DRE, or more typically, a change in or an elevated serum PSA. Genetic risk is also considered.

PHYSICAL EXAMINATION The DRE focuses on prostate size, consistency, and abnormalities within or beyond the gland. Many cancers occur in the peripheral zone and may be palpated on DRE. Carcinomas are characteristically hard, nodular, and irregular, while induration may also be due to benign prostatic hyperplasia (BPH) or calculi. Overall, 20–25% of men with an abnormal DRE have prostate cancer.

PROSTATE-SPECIFIC ANTIGEN PSA (kallikrein-related peptidase 3; *KLK3*) is a kallikrein-related serine protease that causes liquefaction of seminal coagulum. It is produced by both nonmalignant and malignant epithelial cells and, as such, is prostate-specific, not prostate cancer-specific. Serum levels of PSA may increase from prostatitis, BPH, or

prostate cancer. Serum levels are not significantly affected by the DRE. PSA circulating in the blood is inactive and mainly occurs as a complex with the protease inhibitor α_1 -antichymotrypsin and as free (unbound) PSA forms. The formation of complexes between PSA, α_2 -macroglobulin, or other protease inhibitors is less significant. Free PSA is rapidly eliminated from the blood by glomerular filtration with an estimated half-life of 12–18 h. Elimination of PSA bound to α_1 -antichymotrypsin is slow (estimated half-life of 1–2 weeks) as it, too, is largely cleared by the kidneys. Levels should be undetectable after about 6 weeks if the prostate has been completely removed (radical prostatectomy).

PSA testing was approved by the U.S. Food and Drug Administration (FDA) in 1994 for early detection of prostate cancer, and the widespread use of the test has played a significant role in the proportion of men diagnosed with early-stage cancers: >70–80% of newly diagnosed cancers are clinically organ confined. The level of PSA in blood is strongly associated with the risk and outcome of prostate cancer. A single PSA measured at age 60 is associated (area under the curve [AUC] of 0.90) with lifetime risk of death from prostate cancer. Most (90%) prostate cancer deaths occur among men with PSA levels in the top quartile (>2 ng/mL), although only a minority of men with PSA >2 ng/mL will develop lethal prostate cancer. Despite this and mortality rate reductions reported from large randomized prostate cancer screening trials, routine use of the test remains controversial.

In 2012, the U.S. Preventive Services Task Force (USPSTF) published a review of the evidence for PSA-based screening for prostate cancer and made a clear recommendation against screening. By giving a grade of "D" in the recommendation statement that was based on this review, the USPSTF concluded that "there is moderate or high certainty that this service has no net benefit or that the harms outweigh the benefits." In 2013, the American Urological Association (AUA) updated their consensus statement regarding prostate cancer screening. They concluded that the quality of evidence for the benefits of screening was moderate for men aged 55–69 years. For men outside this age range, evidence was lacking for benefit, but the harms of screening, including overdiagnosis and overtreatment, remained. The AUA recommends shared decision-making for men aged 55–69 years considering PSA-based screening, a target age group for whom benefits may outweigh harms. Outside this age range, PSA-based screening as a routine was not recommended. The entire guideline is available at [http://www.auanet.org/guidelines/early-detection-of-prostate-cancer-\(2013-reviewed-and-validity-confirmed-2015\)](http://www.auanet.org/guidelines/early-detection-of-prostate-cancer-(2013-reviewed-and-validity-confirmed-2015).). As of 2017, the USPSTF has issued a revised recommendation with a grade of "C" for PSA-based

prostate cancer screening for men aged 55–69. Now they recommend shared decision-making for men between the ages of 55 and 69 and do not recommend screening for men aged 70 or greater, roughly in agreement with the 2013 AUA guideline. The USPSTF also notes that the increased use of active surveillance (observation with selective delayed treatment) for low-risk prostate cancer has reduced the risks of screening.

We believe that implementation of the following three guidelines will further improve PSA screening outcomes in the United States and will have a greater practical impact on men's health than the USPSTF and AUA recommendations that are based almost solely on age. First, avoid PSA tests in men with little to gain. There is no rationale for recommending PSA screening in asymptomatic men with a short life expectancy. Hence, men over the age of 75 should only be tested in special circumstances, such as a higher than median PSA measured before age 70 or excellent overall health. In addition, because a baseline PSA is a strong predictor of the future risk of lethal prostate cancer, men with low PSAs, for example <1 ng/mL, can undergo testing less frequently, perhaps every 5 years, with screening possibly ending at age 60 if the PSA remains at ≤1 ng/mL. Men with PSAs that are above an age median but below biopsy thresholds can be counseled about their elevated risk and actively encouraged to return for regular screening and more comprehensive risk assessment. Second, do not treat those who do not need treatment. High proportions of men with screen-detected prostate cancer do not need immediate treatment and can be managed by active surveillance. Third, refer men who do need treatment to high-volume centers. Although it is clearly not feasible to restrict treatment exclusively to high-volume centers, shifting treatment trends so that more patients are treated at such centers by high-volume providers will improve cancer control and decrease complications. The goal of prostate cancer screening should be to maximize the benefits of PSA testing and minimize its harms. Following the three rules outlined here should continue to improve the ratio of harms to benefits from PSA screening.

The PSA criteria used to recommend a diagnostic prostate biopsy have evolved over time. However, based on the commonly used cut point for prostate biopsy (a total PSA ≥4 ng/mL), most men with a PSA elevation do not have histologic evidence of prostate cancer at biopsy. In addition, many men with PSA levels below this cut point harbor cancer cells in their prostate. Information from the Prostate Cancer Prevention Trial demonstrates that there is no PSA below which the risk of prostate cancer is zero. Thus, the PSA level establishes the likelihood that a man will harbor cancer if he undergoes a prostate biopsy. The goal is to increase the sensitivity of the test for younger men harboring clinically significant cancers that may cause symptoms and shorten survival and to reduce the frequency of detecting cancers of low malignant potential in elderly men more likely to die of other causes. Patients with symptomatic bacterial prostatitis should have a course of antibiotics before biopsy. However, the routine use of antibiotics in an asymptomatic man with an elevated PSA level is strongly discouraged.

SECOND-LINE SCREENING TESTS Several tests have been developed to better stratify men with an elevated PSA test into those more or less likely to have clinically significant prostate cancer. The 4Kscore® Test (OPKO Lab, Nashville, TN) measures four prostate-specific kallikreins (total PSA, free PSA, intact PSA, and human kallikrein 2). The results are combined with clinical information in an algorithm that estimates an individual's percent risk of having an aggressive prostate cancer should that individual opt for a prostate biopsy. The 4Kscore test has also been shown to identify the likelihood that an individual will develop aggressive prostate cancer, defined as high-grade prostate cancer pathology and/or poor prostate cancer clinical outcomes, within 20 years.

The Prostate Health Index (PHI™, Innovative Diagnostic Laboratory, Richmond, VA) is a blood test that estimates the risk of having prostate cancer. The PHI test is a combination of free PSA, total PSA, and the [-2]proPSA isoform of free PSA. These three tests are combined in a formula that calculates the PHI score. The PHI score is a better predictor of prostate cancer than the total PSA test alone or the free PSA test alone. Urine-based testing measuring exosomes (ExoDx Prostate Test) or mRNA levels of prostate cancer-related genes (Select-MDX) is also available.

PROSTATE BIOPSY A diagnosis of cancer is established by an image-guided needle biopsy. Direct visualization by transrectal ultrasound (TRUS), magnetic resonance imaging (MRI), or fusion of the ultrasound and MRI images ensures that all areas of the gland, including suspicious areas, are sampled. Contemporary schemas advise an extended-pattern 12-core biopsy that includes sampling from the peripheral zone as well as a lesion-directed palpable nodule or suspicious image-guided sampling. Because a prostate biopsy is subject to sampling error, men with an abnormal PSA and negative biopsy are frequently advised to undergo additional testing, which may include a 4Kscore test, PHI, prostate MRI, and/or repeat biopsy.

PATHOLOGY Each core of the biopsy is examined for the presence of cancer, and the amount of cancer is quantified based on the length of the cancer within the core and the percentage of the core involved. Of the cancers identified, >95% are adenocarcinomas; the rest are squamous or transitional cell tumors or, rarely, carcinosarcomas or small-cell histologies. Metastases to the prostate are rare, but in some cases, colon cancers or transitional cell tumors of the bladder invade the gland by direct extension.

When prostate cancer is diagnosed, a measure of histologic aggressiveness is assigned using the *Gleason grading system*, in which the dominant and secondary glandular histologic patterns are scored from 1 (well differentiated) to 5 (undifferentiated) and summed to give a total score of 2–10 for each tumor. The most poorly differentiated area of tumor (i.e., the area with the highest histologic grade) often determines biologic behavior. The presence or absence of perineural invasion and extracapsular spread is also recorded.

Over the years, the Gleason grading system has undergone several changes. Currently, Gleason total scores of 2–5 are no longer assigned, and in practice, the lowest total score is now assigned a 6, although the scale continues to range from 2 to 10. This leads to a logical yet incorrect assumption on the part of patients that their Gleason 6 cancer is in the middle of the scale, triggering the fear that their cancer is serious and the assumption that treatment is necessary despite Gleason score 6 being favorable risk. To address these issues, a new five-grade group system has been developed:

- Grade group 1 (Gleason score ≤6)
- Grade group 2 (Gleason score 3+4 = 7)
- Grade group 3 (Gleason score 4+3 = 7)
- Grade group 4 (Gleason score 4+4 = 8)
- Grade group 5 (Gleason scores 9 and 10)

The new system simplifies the grading of prostate cancer, appropriately classifying the lowest risk as grade group 1 (rather than Gleason score 6), and accurately predicts prognosis.

PROSTATE CANCER STAGING The TNM (tumor, node, metastasis) staging system includes categories for cancers that are identified solely on the basis of an abnormal PSA (T1c), those that are palpable but clinically confined to the gland (T2), and those that have extended outside the gland (T3 and T4) (Table 87-1, Fig. 87-2). DRE alone is inaccurate in determining the extent of disease within the gland, the presence or absence of capsular invasion, involvement of seminal vesicles, and extension of disease to lymph nodes. Because of the inadequacy of DRE for staging, the TNM staging system was modified to include the results of imaging. Unfortunately, no single test has been proven to accurately indicate the stage or the presence of organ-confined disease, seminal vesicle involvement, or lymph node spread.

TRUS is the imaging technique most frequently used to assess the primary tumor, but its chief use is directing prostate biopsies, not staging. No TRUS finding consistently indicates cancer with certainty. Computed tomography (CT) lacks sensitivity and specificity to detect extraprostatic extension and is inferior to MRI in visualization of lymph nodes. In general, MRI is superior to CT to detect cancers in the prostate, to assess local disease extent, and fused with ultrasound imaging, to guide sites to biopsy within the gland. MRI is also useful for the planning of surgery and radiation therapy.

Radionuclide bone scans (bone scintigraphy) are used to evaluate spread to osseous sites. This test is sensitive but relatively nonspecific

TABLE 87-1 TNM Classification

TNM (tumor, node, metastasis) Staging System for Prostate Cancer ^a	
Tx	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Localized Disease	
T1	Clinically inapparent tumor, neither palpable nor visible by imaging
T1a	Tumor incidental histologic finding in ≤5% of resected tissue; not palpable
T1b	Tumor incidental histologic finding in >5% of resected tissue
T1c	Tumor identified by needle biopsy (e.g., because of elevated PSA)
T2	Tumor confined within prostate ^b
T2a	Tumor involves half of one lobe or less
T2b	Tumor involves more than one half of one lobe, not both lobes
T2c	Tumor involves both lobes
Local Extension	
T3	Tumor extends through the prostate capsule ^c
T3a	Extracapsular extension (unilateral or bilateral)
T3b	Tumor invades seminal vesicles
T4	Tumor is fixed or invades adjacent structures other than seminal vesicles such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall
Metastatic Disease	
N1	Positive regional lymph nodes
M1	Distant metastases

^aRevised from SB Edge et al (eds): *AJCC Cancer Staging Manual*, 7th ed. New York, Springer, 2010. ^bTumor found in one or both lobes by needle biopsy, but not palpable or reliably visible by imaging, is classified as T1c. ^cInvasion into the prostatic apex or into (but not beyond) the prostatic capsule is classified not as T3 but as T2.

Abbreviation: PSA, prostate-specific antigen.

because it does not detect the cancer itself, only reaction of the bone to the presence of the cancer itself. Consequently, areas of increased uptake are not always related to metastatic disease. Healing fractures, arthritis, Paget's disease, and other conditions will also cause abnormal uptake. True-positive bone scans are uncommon when the PSA is <10 ng/mL unless the tumor is high-grade.

TREATMENT

Prostate Cancer

CLINICALLY LOCALIZED PROSTATE CANCER

Patients with clinically localized disease are managed by radical prostatectomy, radiation therapy, or active surveillance. Choice of therapy requires the consideration of several factors: the presence of symptoms, the probability that the untreated tumor will adversely affect the quality or duration of survival and thus require treatment,

and the probability that the tumor can be cured by single-modality therapy directed to the prostate versus requiring both local and systemic therapy to achieve cure.

There is no clear evidence for the superiority of any one form of local therapy relative to another. This is due to the lack of prospective randomized trials, referral bias and physician bias, variation in the experience of the treating teams, and differences in trial endpoints and the definitions of cancer control. Often, PSA relapse-free survival is used because an effect on metastatic progression or survival may not be apparent for years. For many patients, however, a PSA recurrence does not necessarily mean that the disease will cause symptoms or shorten survival. After radical surgery to remove all prostate tissue, PSA should become undetectable in the blood within 6 weeks. If PSA remains or becomes detectable after radical prostatectomy, the patient is considered to have persistent or recurrent disease. After radiation therapy, in contrast, PSA does not become undetectable because the remaining nonmalignant elements of the gland continue to produce PSA even if all cancer cells have been eliminated. Similarly, cancer control is not well defined for a patient managed by active surveillance because PSA levels may continue to rise in the absence of therapy. Other outcomes are time to objective progression (local or systemic), cancer-specific survival, and overall survival; however, these outcomes may take years to assess.

The more extensive the local disease, the higher the probability of regional lymph node involvement (even when imaging studies are normal), the lower the probability of local control, and the higher the probability of relapse and the development of metastases. More important is that within the categories of clinical stage T1, T2, and T3 disease are cancers with a range of prognoses. Some T3 tumors are curable with therapy directed solely at the prostate, and some T1 lesions have a high probability of systemic relapse that requires the integration of local and systemic therapy to achieve cure. For T1c cancers, stage alone is inadequate to predict outcome and select treatment; other factors must be considered.

To better assess risk and guide treatment selection, many groups have developed prognostic models or nomograms that use a combination of the initial clinical T stage, biopsy Gleason score, the number of biopsy cores in which cancer is detected, and baseline PSA. Some use discrete cut points (PSA <10 or ≥10 ng/mL; Gleason score of ≤6, 7, or ≥8); others employ nomograms that use PSA and Gleason score as continuous variables. More than 100 nomograms have been reported to predict (1) the probability that a clinically significant cancer is present, (2) disease extent (organ-confined vs non-organ-confined, node-negative or -positive), or (3) the probability of treatment success for specific local therapies using pretreatment variables. Considerable controversy exists over what constitutes "high risk" based on a predicted probability of success or failure. In these situations, nomograms and predictive models can only go so far. Exactly what probability of success or failure would lead a physician to recommend and a patient to seek alternative

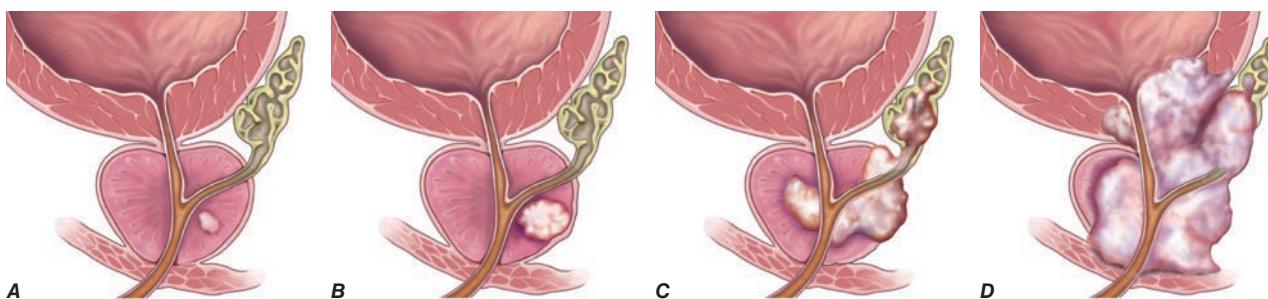


FIGURE 87-2 T stages of prostate cancer. **A.** T1—Clinically inapparent tumor, neither palpable nor visible by imaging. **B.** T2—Tumor confined within prostate. **C.** T3—Tumor extends through prostate capsule and may invade the seminal vesicles. **D.** T4—Tumor is fixed or invades adjacent structures. Eighty percent of patients present with local disease (T1 and T2), which is associated with a 5-year survival rate of 100%. An additional 12% of patients present with regional disease (T3 and T4 without metastases), which is also associated with a 100% survival rate after 5 years. Four percent of patients present with distant disease (T4 with metastases), which is associated with a 30% 5-year survival rate. (Three percent of patients are ungraded.) (Reproduced with permission from MSKCC, data from AJCC, <http://seer.cancer.gov/statfacts/html/prost.html>. © 2010 Memorial Sloan-Kettering Cancer Center Medical Graphics.)

approaches is controversial. As an example, it may be appropriate to recommend radical surgery for a younger patient with a low probability of cure. Nomograms are being refined continually to incorporate additional clinical parameters, biologic determinants, and year of treatment, which can also affect outcomes, making treatment decisions a dynamic process.

Radical Prostatectomy The goal of radical prostatectomy is to excise the cancer completely with a clear margin, to maintain continence by preserving the external sphincter, and to preserve potency by sparing the autonomic nerves in the neurovascular bundle. The procedure is advised for patients with a life expectancy of 10 years or more and is performed via a retropubic or perineal approach or via a minimally invasive robotic-assisted or hand-held laparoscopic approach. Outcomes can be predicted using postoperative nomograms that consider pretreatment factors and the pathologic findings at surgery. PSA failure is usually defined as a value >0.1 or 0.2 ng/mL. Specific criteria to guide the choice of one approach over another are lacking. Minimally invasive approaches offer the advantage of a shorter hospital stay and reduced blood loss. Rates of cancer control, recovery of continence, and recovery of erectile function are comparable. The individual surgeon, rather than the surgical approach used, is most important in determining outcomes after surgery.

Neoadjuvant hormonal treatment with gonadotropin-releasing hormone (GnRH) agonists/antagonists alone has also been explored to improve the outcomes of surgery for high-risk patients using a variety of definitions. The results of several large trials testing 3 or 8 months of androgen depletion before surgery showed that serum PSA levels decreased by 96%, prostate volumes decreased by 34%, and margin positivity rates decreased from 41–17%. Unfortunately, these findings have not been shown to improve PSA relapse-free survival.

Factors associated with incontinence following radical prostatectomy include older age and urethral length, which impacts the ability to preserve the urethra beyond the apex and the distal sphincter. The skill and experience of the surgeon are also factors.

The likelihood of recovery of erectile function is associated with younger age, quality of erections before surgery, and the absence of damage to the neurovascular bundles. In general, erectile function begins to return ~6 months after surgery if neurovascular tissue has been preserved. Potency is reduced by half if at least one neurovascular bundle is sacrificed. Overall, with the availability of drugs such as sildenafil, intraurethral inserts of alprostadil, and intracavernosal injections of vasodilators, many patients recover satisfactory sexual function.

Radiation Therapy Radiation therapy is given by external beam, by radioactive sources implanted into the gland, or by a combination of the two techniques.

External beam radiation therapy Contemporary external beam intensity-modulated radiation therapy (IMRT) permits shaping of the dose and allows the delivery of higher doses to the prostate and a dramatic reduction in normal tissue exposure compared with three-dimensional conformal treatment alone. These advances have enabled the safe administration of doses >80 Gy and resulted in higher local control rates and fewer side effects.

Cancer control after radiation therapy has been defined by various criteria, including a decline in PSA to <0.5 or 1 ng/mL, “nonrising” PSA values, and a negative biopsy of the prostate 2 years after completion of treatment. The current standard definition of biochemical failure (the Phoenix definition) is a rise in PSA by ≥ 2 ng/mL higher than the lowest PSA achieved. Radiation dose is critical to the eradication of prostate cancer. In a representative study, a PSA nadir of <1.0 ng/mL was achieved in 90% of patients receiving 81.0 Gy versus 76% and 56% of those receiving 70.2 and 64.8 Gy, respectively. Positive biopsy rates at 2.5 years were 4% for those treated with 81 Gy versus 27% and 36% for those receiving 75.6 and 70.2 Gy, respectively.

Hypofractionation schedules, utilizing fewer treatments of higher radiation doses, have been evaluated and shown to provide good cancer control rates based on posttreatment biopsies showing no evidence of cancer, with no apparent increase in treatment-related morbidity. Hypofractionated treatments can range from as few as 5 treatments to upward of 26 treatments, both regimens representing substantial reductions in treatment length.

Multiple clinical trials have evaluated the use of androgen deprivation therapy (ADT) in combination with radiation. In patients with unfavorable intermediate-risk prostate cancer, short-course ADT (6 months), when combined with external beam radiotherapy, has demonstrated significant improvements in overall survival. In patients with high-risk disease, longer courses of ADT (18–36 months) have proven superior to shorter courses and represent the current standard of care when combined with radiotherapy.

The Prostate Testing for Cancer and Treatment (ProtecT) trial investigated the effects of active monitoring, radical prostatectomy, and radical radiotherapy with hormones on patient-reported outcomes in men diagnosed with low- and intermediate-risk prostate cancer (~75% with Gleason score 6 or grade group 1 cancer). Patient-reported outcomes among 1643 men who completed questionnaires before diagnosis, at 6 and 12 months, and annually thereafter were compared. Of the three treatments, prostatectomy had the greatest negative effect on sexual function and urinary continence, and although there was some recovery, these outcomes remained worse in the prostatectomy group than in the other groups throughout the trial. The negative effect of radiotherapy on sexual function was greatest at 6 months, but sexual function then recovered somewhat and was stable thereafter; radiotherapy had little effect on urinary continence. Sexual and urinary function declined gradually in the active-monitoring group. Bowel function was worse in the radiotherapy group at 6 months than in the other groups but then recovered somewhat, except for the increasing frequency of bloody stools; bowel function was unchanged in the other groups. Urinary voiding and nocturia were worse in the radiotherapy group at 6 months but then mostly recovered and were like the other groups after 12 months. Effects on quality of life mirrored the reported changes in function. No significant differences were observed among the groups in measures of anxiety, depression, or general health-related or cancer-related quality of life.

Brachytherapy Brachytherapy is the direct implantation of radioactive sources (seeds) into the prostate. It is based on the principle that the deposition of radiation energy in tissues decreases as a function of the square of the distance from the source (Chap. 73). The goal is to deliver intensive irradiation to the prostate, minimizing the exposure of the surrounding tissues. The current standard technique achieves a more homogeneous dose distribution by placing seeds according to a customized template based on imaging assessment of the cancer and computer-optimized dosimetry. The implantation is performed transperineally as an outpatient procedure with real-time imaging.

Improvements in brachytherapy techniques have resulted in fewer complications and a marked reduction in local failure rates. In a series of 197 patients followed for a median of 3 years, 5-year actuarial PSA relapse-free survival for patients with pretherapy PSA levels of 0–4, 4–10, and >10 ng/mL were 98, 90, and 89%, respectively. In a separate report of 201 patients who underwent posttreatment biopsies, 80% were negative, 17% were indeterminate, and 3% were positive. The results did not change with longer follow-up. Brachytherapy is well tolerated, although most patients experience urinary frequency and urgency that can persist for several months. Higher complication rates are observed in patients who have undergone a prior transurethral resection of the prostate (TURP), while those with obstructive symptoms at baseline are at a higher risk for retention and persistent voiding symptoms. Proctitis has been reported in <2% of patients.

Active surveillance With the advent of PSA testing, many patients are diagnosed with low-risk prostate cancers that may

not pose a threat to either the quantity or quality of man's life. Active surveillance, described previously as *watchful waiting* or *deferred therapy*, evolved from (1) studies that evaluated predominantly elderly men with well-differentiated tumors who remained untreated and demonstrated no clinically significant progression for protracted periods, (2) recognition of the contrast between incidence and disease-specific mortality, (3) the high prevalence of autopsy cancers, and (4) an effort to reduce overtreatment and treatment-related side effects. In practice, active surveillance is the treatment recommended to patients with cancers of low aggressiveness that can be safely monitored at fixed intervals with DREs, PSA measurements, imaging (usually prostate MRI), and repeat prostate biopsies as indicated until histopathologic or serologic changes correlative of progression warrant treatment with curative intent.

Case selection is critical, and determining the clinical parameters predictive of cancer aggressiveness that can be used to reliably select men most likely to benefit from active surveillance is an area of intense study. One set of criteria includes men with clinical T1c tumors that are biopsy Gleason grade 6 (grade group 1) involving three or fewer cores, with each core having <50% involvement by tumor, and a PSA density of <0.15. Nomograms to help predict which patients can safely be managed by active surveillance continue to be refined, and as their predictive accuracy improves, it can be anticipated that more patients will be candidates.

RISING PSA AFTER DEFINITIVE LOCAL THERAPY

Patients in this state include those in whom the sole manifestation of disease is a rising PSA after surgery and/or radiation therapy. There is no evidence of disease on imaging studies. For these patients, the central issue is whether the rise in PSA results from persistent disease in the primary site, systemic disease, or both. In theory, disease in the primary site may still be curable by additional local treatment.

The decision to recommend radiation therapy after prostatectomy is guided by the pathologic findings at surgery, the timing of PSA failure, and the PSA level at the time of failure. Traditional imaging (MRI, CT, and radionuclide bone scans), especially at low levels of PSA, are typically uninformative. New positron emission tomography (PET) tracers such as C-11 choline, F-18 fluciclovine, and F-18 or Ga-68 prostate-specific membrane antigen (PSMA) that directly image the cancer are more sensitive and can detect low-volume disease in the prostate bed or other sites to better inform the decision to recommend additional local therapies. All are FDA approved. Detection rates, both in and outside the prostate bed, correlate with the absolute level of PSA. Factors that predict for response to salvage radiation therapy are a positive surgical margin, lower Gleason score in the radical prostatectomy specimen, long interval from surgery to PSA failure, slow PSA doubling time, and low (<0.5 ng/mL) PSA value at the time of radiation treatment. For patients with a rising PSA after radiation therapy, salvage local therapy can be considered if the disease was "curable" at the outset, if persistent disease has been documented by a biopsy of the prostate, and if no disease is detectable outside of the prostate bed or regional lymph nodes by imaging. Unfortunately, case selection is poorly defined in most series, and morbidities are significant. Options include salvage radical prostatectomy, salvage cryotherapy, salvage radiation therapy, and salvage high-intensity focused ultrasound.

The rise in PSA after surgery or radiation therapy may indicate subclinical or micrometastatic disease with or without local recurrence. In these cases, the need for treatment depends, in part, on the estimated probability that the patient will show evidence of metastatic disease on a scan and in what time frame. That immediate therapy is not always required was shown in a series where patients who developed a rising PSA after radical prostatectomy received no systemic therapy until metastatic disease was documented. Overall, the median time to metastatic progression was 8 years, and 63% of the patients with rising PSA values remained free of metastases at 5 years. Factors associated with progression included the Gleason score of the radical prostatectomy specimen, time to recurrence after surgery, and PSA doubling time. For those with Gleason

score ≥8, the probability of metastatic progression was 37, 51, and 71% at 3, 5, and 7 years, respectively. If the time to recurrence was <2 years and PSA doubling time was long (>10 months), the proportions with metastatic disease at the same time intervals were 23, 32, and 53%, versus 47, 69, and 79% if the doubling time was short (<10 months). PSA doubling times are also prognostic for survival. In one series, all patients who succumbed to disease had PSA doubling times of ≤3 months. Most physicians advise treatment when PSA doubling times are ≤12 months. A difficulty with predicting the risk of metastatic spread, symptoms, or death from disease in the rising PSA state is that most patients receive some form of therapy before the development of metastases. Nevertheless, predictive models continue to be refined.

METASTATIC DISEASE: NONCASTRATE

The state of *noncastrate metastatic disease* includes men with metastases visible on an imaging study at the time of diagnosis or after local therapy(ies) who have testosterone levels >150 ng/dL. Symptoms of metastatic disease include pain from osseous spread, although many patients are asymptomatic despite extensive spread. Less common are symptoms related to marrow infiltration by tumor (myelophthisis), coagulopathy, or spinal cord compression. Standard treatment is to deplete or lower androgens via ADT by medical or surgical means, the latter being the least acceptable to patients. A less frequently used approach is to block androgen binding to the AR with antiandrogens. More than 90% of male hormones originate in the testes; <10% are synthesized in the adrenal gland (Fig. 87-3).

Testosterone-Lowering Agents Medical therapies that lower testosterone levels include the GnRH agonists/antagonists, pure GnRH antagonists, 17,20-lyase inhibitors, CYP17 inhibitors, and estrogens such as diethylstilbestrol (DES). The latter are rarely utilized due to the risk of vascular complications that include fluid retention, phlebitis, emboli, and stroke. GnRH agonists/antagonists, such as leuprolide acetate and goserelin acetate, initially produce a rise in luteinizing hormone and follicle-stimulating hormone followed by a downregulation of receptors in the pituitary gland, which effects a chemical castration. Regulatory approval was based on randomized trials showing reduced cardiovascular toxicities relative to DES, with equivalent potency. The initial rise in testosterone may result in a clinical flare of the disease, and as such, these agents are relatively contraindicated in men with significant obstructive symptoms, cancer-related pain, or spinal cord compromise, events that do not occur with GnRH antagonists such as degarelix, given by injection, or relugolix, given orally, that rapidly achieve castrate levels of testosterone. AR antagonists that block testosterone binding to the receptor are also used to prevent flare.

Agents that lower testosterone are associated with an androgen-deprivation syndrome that includes hot flushes, weakness, fatigue, loss of muscle mass, anemia, changes in cognition and personality, and depression. Changes in lipids, obesity, insulin resistance, and an increased risk of diabetes and cardiovascular disease are also seen, along with a decrease in bone density that worsens over time and results in an increased risk of clinical fractures. This is a particular concern in men with preexisting osteopenia that results from hypogonadism that may be worsened with steroid or alcohol use and significantly underappreciated. Baseline fracture risk can be assessed using the FRAX scale, and to minimize fracture risk, patients are advised to take calcium and vitamin D supplementation, along with a bisphosphonate, RANK-ligand inhibitor (denosumab), or toremifene.

Antiandrogens Nonsteroidal first-generation antiandrogens such as bicalutamide and nilutamide have largely been replaced by the more potent next-generation agents (enzalutamide, apalutamide, and darolutamide) that do not lower serum androgen levels and result in fewer hot flushes, less of an effect on libido, less muscle wasting, fewer personality changes, and less bone loss relative to testosterone-lowering therapies. However, over time, testosterone

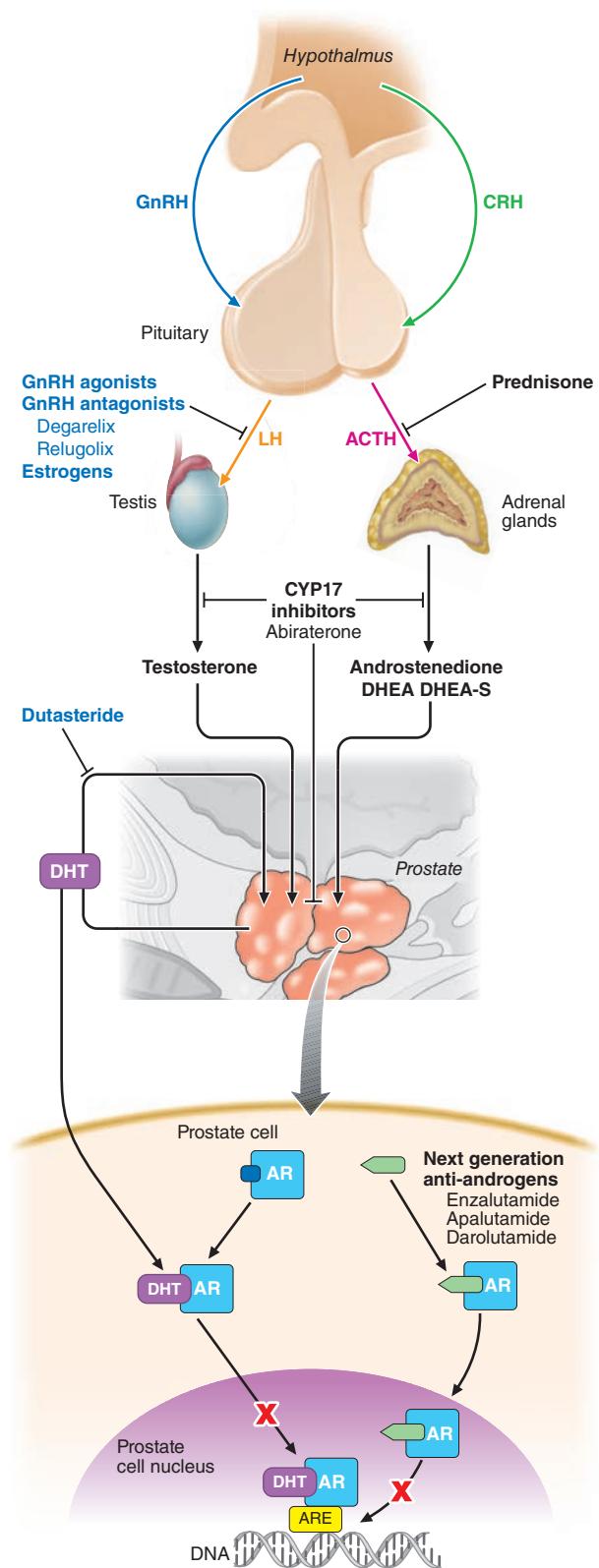


FIGURE 87-3 Sites of action of different hormone therapies.

levels increase and are converted to estrogen, which can result in mastalgia and gynecomastia that limits long-term use but can be prevented in part by tamoxifen or prophylactic breast irradiation.

Most reported randomized trials suggest that the cancer-specific outcomes are inferior when antiandrogens are used alone. Bicalutamide, even at a dose of 150 mg (three times the approved dose for use in combination in GnRH agonists), resulted in a shorter time to progression and inferior survival compared with surgical castration for patients with established metastatic disease.

Improving on the outcomes with ADT alone was a focus of the field for decades. One approach was to combine a first-generation antiandrogen (flutamide, bicalutamide, or nilutamide) with a GnRH analogue or surgical orchietomy; however, this approach did not improve outcomes, and current use is largely limited to the first 2–4 weeks of treatment to protect against flare.

Practice standards changed when an improvement in time to progression and overall survival was shown when ADT was combined with docetaxel, the first systemic therapy shown to prolong life in metastatic castration-resistant prostate cancer (mCRPC) approved in 2004, relative to ADT alone. The greatest benefit was seen for patients with “high-volume” disease defined as the presence of ≥4 lesions on radionuclide bone scan or visceral disease. For abiraterone acetate and prednisone, benefit was seen across disease states ranging from high-risk localized to metastatic disease. Longer progression-free and overall survival times have been noted in separate phase 3 trials comparing ADT with abiraterone, a CYP17 inhibitor that blocks androgen synthesis, and ADT with the AR antagonists enzalutamide and apalutamide versus the ADT standard, further changing the standards of care.

Intermittent Androgen Deprivation Therapy (IADT) One way to reduce the side effects of androgen depletion is to administer antiandrogens on an intermittent basis. This was proposed as a way to prevent the selection of cells that are resistant to androgen depletion. The hypothesis is that by allowing endogenous testosterone levels to rise, the cells that survive androgen depletion will induce a normal differentiation pathway. In this way, the surviving cells that are allowed to proliferate in the presence of androgen will retain sensitivity to subsequent androgen depletion. Applied in the clinic, androgen depletion is continued for 2–6 months beyond the point of maximal response. Once treatment is stopped, endogenous testosterone levels increase, and the symptoms associated with hormone treatment abate. PSA levels also begin to rise, and at some level, treatment is restarted. With this approach, multiple cycles of regression and proliferation have been documented in individual patients. Unknown is whether the intermittent approach increases, decreases, or does not change the overall duration of sensitivity to androgen depletion. The approach is safe, but long-term data are needed to assess the course in men with low PSA levels. A trial to address this question is ongoing.

Outcomes of Androgen Deprivation The anti-prostate cancer effects of the various androgen depletion strategies are similar, and the clinical course is predictable: an initial response, a period of stability in which tumor cells are dormant and nonproliferative, followed after a variable period of time by a rise in PSA and regrowth that is visible on a scan as a castration-resistant lesion. Androgen depletion is not curative because cells that survive castration are present when the disease is first diagnosed. Considered by disease manifestation, PSA levels return to normal in 60–70% of patients, and measurable disease regression occurs in 50%; improvements in bone scan occur in 25% of cases, but the majority remain stable. Duration of survival is inversely proportional to disease extent at the time androgen depletion is first started and the nadir level of PSA at 6 months. Patients with nadir values above a certain threshold have markedly inferior survival times and should be considered for alternative approaches.

An unresolved question remains on how early systemic therapies should be offered to patients: in the adjuvant setting after surgery or radiation treatment of the primary tumor; at the time that a PSA recurrence is documented; or wait until metastatic disease or symptoms of disease are manifest? Trials in support of early therapy have been largely underpowered relative to the reported benefit or have been criticized on methodologic grounds. One that showed a survival benefit for patients treated with radiation therapy and 3 years of ADT, relative to radiation alone, was criticized for the poor outcomes of the control group. Another showing a survival benefit for patients with positive lymph nodes who were randomized to immediate medical or surgical castration compared with

observation ($p = .02$) was criticized because the confidence intervals around the 5- and 8-year survival distributions for the two groups overlapped.

METASTATIC DISEASE: CASTRATE

Castration-resistant prostate cancer (CRPC), disease that progresses while the measured levels of testosterone in the blood are 50 ng/mL or lower, can produce some of the most feared complications of the disease and is lethal for most men. The most common manifestation is a rising PSA, frequently co-occurring with progression in bone. Nodal and/or visceral spread is less frequent. Symptoms may or may not be present. The bone- and PSA-dominant pattern limits the ability to assess treatment effects reliably because traditional bone imaging is inaccurate and no PSA-based outcome has been shown to be a true surrogate for survival, and accordingly, favorable changes with either can be used to support regulatory approvals. Critical for management is that therapeutic objectives be based on the manifestations of the disease in the individual at the time a change in therapy is being considered. As such, for the patient with symptomatic bone disease, relief of pain can be more clinically relevant than lowering the PSA. Naturally, for all patients, the central focus is delaying or preventing disease progression, symptoms, and death from disease.

Through 2010, docetaxel was the only FDA-approved life-prolonging therapy for CRPC. Since then, our understanding of the biology of the disease has increased significantly, which in turn has led to improved therapies. In particular, it is now recognized that the majority of mCRPCs continue to express the AR and remain AR signaling dependent, and upward of 50% of cases harbor a series of oncogenic changes including overexpression, splice variants lacking the ligand binding domain and that stimulate growth independent of the ligand, and upregulation of the enzymes in the androgen biosynthesis pathway, leading to an increase in intratumoral androgens. These oncogenic changes have been successfully targeted with the next-generation antiandrogens enzalutamide, apalutamide, and darolutamide and the CYP17 inhibitor abiraterone acetate (given in combination with prednisone), all of which have been proven to prolong life and are FDA approved for use in CRPC in both the pre- and postchemotherapy setting.

Large-scale molecular profiling efforts have led to a biologically based disease taxonomy that continues to evolve and showed a markedly higher than expected frequency of germline and somatic *BRCA2* alterations, along with other genes in the DNA damage repair pathway that have been targeted successfully with poly-ADP ribose polymerase (PARP) inhibitors of which two, olaparib and rucaparib, are FDA approved, and one, niraparib, has achieved a breakthrough designation. Also approved is the checkpoint inhibitor pembrolizumab for tumors with high microsatellite instability (MSI) scores, an alteration found in 2–3% of prostate cancers for which a dedicated prostate cancer trial would never have been conducted.

Other classes of therapy are approved based on a demonstrated survival benefit include the biologic agent sipuleucel-T, the second-generation taxane cabazitaxel, and the α -emitting bone-targeting radiopharmaceutical radium-223. Approval is also anticipated for PSMA-directed radionuclide therapy based on the survival benefit of Lu-177 PSMA in the phase 3 VISION trial relative to best supportive care alone. Overall, an intense focus of current CRPC research is to understand the optimal sequence in which to utilize these agents to maximize benefit for the individual patient. Most of these agents are also being tested earlier in the course of the disease when tumor burdens are lower and the disease less heterogeneous. The result has been an increase in the frequency of late-state tumors that have undergone a lineage transformation from epithelial to neuroendocrine phenotypes and are highly resistant to available therapies.

Pain Management Pain secondary to osseous metastases is one of the most feared complications of the disease and a major cause of morbidity, worsened by the narcotics needed to control symptoms.

Management requires accurate diagnoses because noncancer etiologies including degenerative disease, spinal stenosis, and vertebral collapse secondary to bone loss are common. Neurologic symptoms, including those suggestive of base of skull disease or spinal cord compromise, require emergency evaluation because loss of function may be permanent if not addressed quickly. Neurologic symptoms and loss of function are best treated with external beam radiation, as are single sites of pain. Diffuse symptoms in the absence of neurologic deficits can be treated with bone-seeking radioisotopes, such as radium-223 or the β emitter ^{153}Sm -EDTMP; mitoxantrone; or other systemic therapies, such as abiraterone acetate, enzalutamide, and docetaxel. Radium-223 is indicated for patients with symptoms, whereas ^{153}Sm -EDTMP and mitoxantrone are approved for the palliation of pain but have not been shown to prolong life. Abiraterone, enzalutamide, and docetaxel do not have a formal indication for pain but were shown to palliate pain in the registration trials that led to their approval by showing a survival benefit.

Other bone-targeting agents, including bisphosphonates such as zoledronic acid and the RANK-ligand inhibitor denosumab, have been shown to reduce the frequency and development of skeletal complications including pain requiring analgesia, neurologic compromise from epidural extension of tumor, and/or the need for surgery or radiation therapy to treat symptomatic osseous disease. It is important to note that, for all of these agents, the direct effect on the tumor is modest, and benefits are seen without declines in PSA or improvements on imaging.

BENIGN DISEASE

■ BENIGN PROSTATIC HYPERPLASIA

BPH is a pathologic process that contributes to the development of lower urinary tract symptoms (LUTS) in men. LUTS, arising from lower urinary tract dysfunction, are further subdivided into obstructive symptoms (urinary hesitancy, straining, weak stream, terminal dribbling, prolonged voiding, incomplete emptying) and irritative symptoms (urinary frequency, urgency, urge incontinence, small voided volumes). LUTS and other sequelae of BPH are not just due to a mass effect but are also likely due to a combination of the prostatic enlargement and age-related detrusor dysfunction.

Diagnostic Procedures and Treatment LUTS are generally measured using a validated, reproducible index that is designed to determine disease severity and response to therapy—the AUA's Symptom Index (AUASI), also adopted as the International Prostate Symptom Score (IPSS) (Table 87-2). Serial AUASI is particularly useful in following patients as they are treated with various forms of therapy. Asymptomatic patients do not require treatment regardless of the size of the gland, while those with an inability to urinate, gross hematuria, recurrent infection, or bladder stones require evaluation and treatment. In patients with symptoms, uroflowmetry can identify those with normal flow rates who are unlikely to benefit from treatment, and bladder ultrasound can identify those with high postvoid residuals who may need intervention. Pressure-flow (urodynamic) studies detect primary bladder dysfunction. Cystoscopy is recommended if hematuria is documented and to assess the urinary outflow tract before surgery. Imaging of the upper tracts is advised for patients with hematuria, a history of calculi, or prior urinary tract problems.

Symptomatic relief is the most common reason men seek treatment for BPH, and therefore, symptomatic relief is usually the goal of therapy for BPH. α -Adrenergic receptor antagonists are thought to treat the dynamic aspect of BPH by reducing sympathetic tone of the bladder outlet, thereby decreasing resistance and improving urinary flow. 5ARIs are thought to treat the static aspect of BPH by reducing prostate volume and having a similar, albeit delayed effect. 5ARIs have also proven beneficial in the prevention of BPH progression, as measured by prostate volume, the risk of developing acute urinary retention, and the risk of having BPH-related surgery. The use of an alpha-adrenergic

TABLE 87-2 AUA Symptom Index

QUESTIONS TO BE ANSWERED	AUA SYMPTOM SCORE (CIRCLE 1 NUMBER ON EACH LINE)					
	NOT AT ALL	LESS THAN 1 TIME IN 5	LESS THAN HALF THE TIME	ABOUT HALF THE TIME	MORE THAN HALF THE TIME	ALMOST ALWAYS
Over the past month, how often have you had a sensation of not emptying your bladder completely after you finished urinating?	0+	1	2	3	4	5
Over the past month, how often have you had to urinate again less than 2 h after you finished urinating?	0	1	2	3	4	5
Over the past month, how often have you found you stopped and started again several times when you urinated?	0	1	2	3	4	5
Over the past month, how often have you found it difficult to postpone urination?	0	1	2	3	4	5
Over the past month, how often have you had a weak urinary stream?	0	1	2	3	4	5
Over the past month, how often have you had to push or strain to begin urination?	0	1	2	3	4	5
Over the past month, how many times did you most typically get up to urinate from the time you went to bed at night until the time you got up in the morning?	(None)	(1 time)	(2 times)	(3 times)	(4 times)	(5 times)
Sum of 7 circled numbers (AUA Symptom Score): _____						

Abbreviation: AUA, American Urological Association.

Source: Reproduced with permission from MJ Barry et al: The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol* 148:1549, 1992.

receptor antagonist and a 5ARI as combination therapy seeks to provide symptomatic relief while preventing progression of BPH.

Another class of medications that has shown improvement in LUTS secondary to BPH is phosphodiesterase-5 (PDE5) inhibitors, used currently in the treatment of erectile dysfunction. All four of the PDE5 inhibitors available in the United States—sildenafil, vardenafil, tadalafil, and avanafil—appear to be effective in the treatment of LUTS secondary to BPH. The use of PDE5 inhibitors is not without controversy, however, given the fact that short-acting phosphodiesterase inhibitors such as sildenafil need to be dosed separately from alpha blockers such as tamsulosin because of potential hypotensive effects.

Symptoms due to BPH often coexist with symptoms due to overactive bladder, and the most common pharmacologic agents for the treatment of overactive bladder symptoms are anticholinergics. This has led to multiple studies evaluating the efficacy of anticholinergics for the treatment of LUTS secondary to BPH.

Surgical therapy is now considered second-line therapy and is usually reserved for patients after a trial of medical therapy. The goal of surgical therapy is to reduce the size of the prostate, effectively reducing resistance to urine flow. Surgical approaches include TURP, transurethral incision, or removal of the gland via a retropubic, suprapubic, or perineal approach. Also used are transurethral ultrasound-guided laser-induced prostatectomy (TULIP), stents, and hyperthermia.

FURTHER READING

BARRY MJ, SIMMONS LH: Prevention of prostate cancer morbidity and mortality: Primary prevention and early detection. *Med Clin North Am* 101:787, 2017.

BATTAGLIA A et al: Novel insights into the management of oligometastatic prostate cancer: A comprehensive review. *Eur Urol Oncol* 2:174, 2019.

BUYYOUNOUSKI MK et al: Prostate cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 67:245, 2017.

CALAIS J et al: ¹⁸F-fluciclovine PET-CT and ⁶⁸Ga-PSMA-11 PET-CT in patients with early biochemical recurrence after prostatectomy: A prospective, single-centre, single-arm, comparative imaging trial. *Lancet Oncol* 20:1286, 2019.

DE VRIES KC et al: Hypofractionated versus conventionally fractionated radiation therapy for patients with intermediate- or high-risk, localized, prostate cancer: 7-year outcomes from the randomized, multicenter, open-label, phase 3 HYPRO trial. *Int J Radiat Oncol Biol Phys* 106:108, 2020.

HUSSAIN M et al: Survival with olaparib in metastatic castration-resistant prostate cancer. *N Engl J Med* 383:2345, 2020.

JAIRATH NK et al: A systematic review of the evidence for the decipher genomic classifier in prostate cancer. *Eur Urol* 79:374, 2021.

MERSEBURGER AS et al: Genomic testing in patients with metastatic castration-resistant prostate cancer: A pragmatic guide for clinicians. *Eur Urol* 79:519, 2021.

SHORE ND et al: Oral relugolix for androgen-deprivation therapy in advanced prostate cancer. *N Engl J Med* 382:2187, 2020.

VIRGO KS et al: Initial management of noncastrate advanced, recurrent, or metastatic prostate cancer: ASCO guideline update. *J Clin Oncol* 39:1274, 2021.

YAMADA Y et al: Clinical and biological features of neuroendocrine prostate cancer. *Curr Oncol Rep* 23:15, 2021.

88

Testicular Cancer

David J. Vaughn



Testicular germ cell tumors (GCTs) represent 95% of all testicular neoplasms. Non-GCTs of the testis are much less common. Approximately 5% of GCTs arise in extragonadal locations including the mediastinum, retroperitoneum, and pineal gland. Treatment for testicular GCTs is determined by pathology and stage. The development of effective chemotherapy for this disease represents a landmark achievement in oncology. About 95% of newly diagnosed patients with testicular GCTs will be cured. For this reason, testicular cancer has been called “a model for a curable neoplasm.”

INCIDENCE

In 2021, ~9500 cases of testicular GCTs will be diagnosed in the United States, with <450 deaths. These tumors are diagnosed most commonly in men between 20 and 40 years. The incidence of GCTs is increasing in men age 50 years and older.

GLOBAL CONSIDERATIONS

The incidence of testicular GCTs appears to be increasing worldwide. The disease has the highest incidence in Scandinavia, Western Europe, and Australia/New Zealand. Africa and Asia have the lowest incidence. The incidence in the United States and the United Kingdom is intermediate. While there does not appear to be a distinct biology related to geography, several countries have reported a migration to earlier stage disease in part related to public awareness and earlier diagnosis.

EPIDEMIOLOGY

GCTs are predominantly seen in young Caucasian men. The disease is much less commonly seen in African Americans. Testicular GCTs have an estimated heritability of almost 50%. Interestingly, the risk of GCT is higher in male siblings than in offspring of the patient. Although epidemiologic studies have been performed attempting to identify a relationship with environmental exposures, no conclusive causal links have been established.

Risk Factors The strongest risk factors for testicular GCT include a prior history of the disease, cryptorchidism, and a history of testicular germ cell neoplasia in situ (GCNIS). Patients with a prior history of testicular GCT have a 2% risk of developing a contralateral GCT. These are more commonly metachronous than synchronous. Men with cryptorchidism have approximately a four- to sixfold increased risk of developing testicular GCT. Orchidopexy before puberty decreases but does not eliminate this risk. Interestingly, the contralateral descended testis is also at risk for this disease. Men undergoing infertility evaluation in which a testicular biopsy demonstrates GCNIS have a significant risk of developing GCT. Although scrotal ultrasound of patients with testicular GCT may demonstrate testicular microcalcifications that may be related to GCNIS, the significance of testicular microcalcifications in the general population is unclear.

BIOLOGY

The primordial germ cell is the cell of origin for GCTs. Most malignant GCTs arise from GCNIS. The molecular events that result in the development of GCNIS and subsequent malignant GCT have not been fully determined. However, genetic analysis of GCTs has demonstrated an excess copy number of isochromosome 12p (i[12p]) in most cases. Several genome-wide association studies have identified multiple independent loci associated with testicular GCT risk. The strongest of these is the *KITLG* (KIT ligand) locus on chromosome 12. These loci contribute significantly to the heritable risk of this disease.

PATHOLOGY

GCTs are either seminomas or nonseminomas. For a tumor to be considered a seminoma, it must be 100% seminoma. Any mixed GCT is best approached as a nonseminomatous GCT (NSGCT). Seminomas represent ~50% of cases. Seminomas arise most commonly in patients in the fourth decade of life. Seminomas may contain syncytiotrophoblastic cells, which may secrete β human chorionic gonadotropin (hCG). Seminomas do not secrete α fetoprotein (AFP). Seminomas are exquisitely sensitive to both chemotherapy and radiation therapy. NSGCTs are most commonly diagnosed in the third decade of life. The histologic subtypes include embryonal carcinoma, yolk sac tumor, choriocarcinoma, and teratoma. Embryonal carcinoma is the most undifferentiated NSGCT subtype with the potential to differentiate into the other subtypes. Embryonal carcinoma may secrete AFP, hCG, both, or neither. Yolk sac tumor often secretes AFP. Choriocarcinoma is an aggressive subtype, often secreting hCG at very high levels. These NSGCT subtypes are all considered chemotherapy sensitive. Teratoma is composed of somatic cell types that are derived from two or more germinal layers (endoderm, mesoderm, and ectoderm). Teratomas are classified as mature, in which cell types resemble normal adult somatic

tissue; immature, in which cell types resemble fetal somatic tissue; and malignant, in which the cell types have undergone malignant transformation into the malignant counterpart of the somatic tissue. Teratomas are chemotherapy resistant and must be approached surgically.

INITIAL PRESENTATION

Signs and Symptoms Although a painless testicular mass is pathognomonic of a GCT, most patients present with testicular swelling, firmness, discomfort, or a combination of these. The differential diagnosis may include epididymitis or orchitis and a trial of antibacterials may be considered. Patients with retroperitoneal metastases may complain of back or flank pain. Patients may have cough, shortness of breath, or hemoptysis as a result of lung metastases. In patients with elevation of serum hCG, gynecomastia may be present. Diagnostic delay is not uncommon and may be associated with a more advanced stage at diagnosis.

Physical Examination Careful examination of the affected testis and the contralateral normal testis should be performed. Many tumors will have a hard consistency to palpation. Some patients may show testicular atrophy. Evaluation for supraclavicular lymphadenopathy, gynecomastia, and abdominal mass should be performed. Inguinal lymphadenopathy is rare. Most patients with lung metastases will have normal auscultation of the lungs.

Diagnostic Testing If a firm testicular mass is identified, a scrotal ultrasound should be performed. Patients with suspected epididymitis or orchitis who do not respond to antibiotics should also undergo scrotal ultrasound. Scrotal ultrasound should include both testicles. On ultrasound, a testicular GCT is hypoechoic and may be multifocal. A solid mass identified on ultrasound should be considered malignant until otherwise proven. Transscrotal aspiration or biopsy of a testicular mass should never be performed. Such scrotal violation may result in tumor seeding of the scrotum or inguinal lymph nodes.

Serum Tumor Markers Serum AFP, hCG, and lactate dehydrogenase (LDH) should be measured in patients suspected of testicular GCT. AFP is elevated in ~60–70% of patients who present with NSGCTs. Seminomas never secrete AFP. A patient with a seminoma with elevation of AFP should be approached as having an NSGCT. The half-life of AFP is 5–7 days. A falsely elevated AFP may be seen in patients with hepatic disease or a condition called hereditary persistence of AFP, in which patients may have baseline AFP levels that are mildly elevated. hCG may be elevated in both NSGCTs as well as seminomas. Patients with choriocarcinoma may have markedly elevated levels of hCG. The half-life for hCG is 24–36 h. False-positive elevation of hCG may be seen secondary to hypogonadism, marijuana use, or as a result of interfering substances measured by the assay. LDH is a non-specific marker for GCT. Its principal use is to help in the assessment of the risk classification of a patient with metastatic disease. Although elevation of serum tumor markers supports the diagnosis of a testicular GCT, it should be remembered that most patients with seminoma and up to a third of patients with NSGCTs do not have elevated levels. Serum microRNA (miR)-371a-3 has been identified as a promising biomarker for GCT, and validation studies are ongoing.

INITIAL MANAGEMENT

Inguinal Orchiectomy Prompt referral to urology should be performed if a testicular GCT is suspected. The initial treatment for most patients suspected of having a testicular GCT is radical inguinal orchiectomy with removal of the testicle and spermatic cord to the level of the internal inguinal ring. In patients who present with metastatic disease and the diagnosis of GCT is certain, orchiectomy may be deferred until completion of chemotherapy. Although some institutions perform testis-sparing surgery in select patients, the gold standard remains radical inguinal orchiectomy. Pathologic examination of the entire testicle is important, since testicular GCTs may be multifocal. Given the rarity of this cancer, review by an experienced pathologist is essential for accurate tumor classification. Serum tumor markers should be obtained before and after orchiectomy.

Staging The staging of testicular GCT is based on an understanding of the pattern of spread. The initial spread is by the lymphatic route to the retroperitoneal lymph nodes. A left-sided testicular GCT spreads first to the primary landing zone of left paraaortic lymph nodes inferior to the left renal vessels. A right-sided testicular GCT spreads first to the primary landing zone of the aortocaval nodes inferior to the right renal vessels. Nodal metastases may extend into the iliac regions. If scrotal violation occurred, inguinal lymph node metastases may be seen. Subsequent lymphatic spread is to the retrocrural, mediastinal, and supraclavicular lymph nodes. Hematogenous spread to the lung is the next most common site of metastasis. Metastases to the liver, bone, and brain are less commonly seen. Patients with newly diagnosed testicular GCTs should undergo computed tomography (CT) scan of the abdomen and pelvis. Chest x-ray should be performed. CT scan of the chest is performed if retroperitoneal metastases are present or if lung nodules are identified on chest x-ray. Bone scan and magnetic resonance imaging (MRI) of the brain are not routinely performed unless clinically indicated. Positron emission tomography (PET) has little role in the initial staging of testicular GCTs.

The American Joint Committee on Cancer tumor-node-metastasis (TNM) staging classification is used. There are three main stages of testicular GCT. Stage I is limited to the testis; stage II involves the retroperitoneal lymph nodes; and stage III includes lymph node involvement beyond the retroperitoneum and/or distant metastatic disease.

■ STAGE-BASED MANAGEMENT

Treatment of testicular GCT is based on two factors: (1) whether the tumor is seminoma or NSGCT and (2) the stage of the patient. This is summarized in Fig. 88-1.

Stage I • SEMINOMA About 70% of newly diagnosed patients with seminoma present with stage I disease. This is defined as no evidence of metastatic disease on imaging of the chest, abdomen, and pelvis. Approximately 15% of patients with stage I seminoma have metastatic disease at the microscopic level, usually in the retroperitoneum. Historically, patients with stage I seminoma were treated with a course of adjuvant radiation therapy to the paraaortic lymph nodes. While still an option, this is not usually performed because of concerns for late radiation-induced secondary malignancies. Active surveillance is the most common approach elected by these patients following orchietomy. With active surveillance, interval physical examination and CT scan of the abdomen are performed. For the 15% of patients who develop metastatic disease during active surveillance, treatment with definitive radiation therapy or chemotherapy is curative in nearly all. A third option for clinical stage I seminoma is adjuvant chemotherapy with carboplatin monotherapy for one or two cycles. While effective in decreasing the risk of recurrence, it should be remembered that most patients are cured by orchietomy alone, and therefore, the additional treatment is unnecessary. In addition, long-term data on toxicity are not available.

NSGCTs About 40% of newly diagnosed patients with NSGCTs present with stage I disease. Because NSGCTs have an increased potential for invasion and metastasis, spread to the retroperitoneum and beyond is more common than with seminoma. If pre-orchietomy serum tumor markers are elevated, these must normalize after orchietomy to be considered stage I. Patients with persistently elevated or rising serum tumor markers after orchietomy have stage IS disease and should be treated with cisplatin-based chemotherapy. If the tumor is limited to testis without lymphovascular invasion, the risk of recurrence is approximately 20%. However, if the tumor has high-risk features including lymphovascular invasion, invasion of the spermatic cord, or invasion of the scrotum, the risk of recurrence is ~50% or higher. Historically, a prophylactic retroperitoneal lymph node dissection (RPLND) was performed. This surgery is not only diagnostic but also therapeutic. In fact, most patients who undergo prophylactic RPLND will never require chemotherapy. While still an option, this approach subjects many patients to unnecessary major abdominal surgery. RPLND is also associated with a small risk of retrograde ejaculation due to nerve injury, and nerve-sparing techniques have been developed. Active surveillance is frequently performed especially for

patients without lymphovascular invasion. Most patients who relapse will be treated with cisplatin-based chemotherapy and achieve cure rates approaching 100%. Active surveillance can also be employed for patients with higher risk features, although the risk of progression is significantly higher. For this reason, some advocate adjuvant cisplatin-based chemotherapy with BEP (bleomycin, etoposide, cisplatin) for one cycle for these patients. Other centers favor a prophylactic RPLND. Almost all patients who present with stage I NSGCTs will achieve cure.

Stage II • SEMINOMA Approximately 15–20% of newly diagnosed patients with seminoma present with stage II disease. Patients are subgrouped into IIA, IIB, or IIC based on the size of the retroperitoneal nodes (≤ 2 cm, >2 to 5 cm, or >5 cm, respectively). Patients with stage IIA disease are usually treated with “dogleg” radiation therapy (referring to the shape of the radiation field), which includes the paraaortic and ipsilateral iliac nodes. Cisplatin-based chemotherapy may also be considered. Stage IIB disease is treated with cisplatin-based chemotherapy or, in select patients, radiation therapy. Most patients treated with radiation therapy who relapse will subsequently be cured with cisplatin-based chemotherapy. For patients with stage IIC disease, cisplatin-based chemotherapy should be used.

NSGCTs Approximately 15% of newly diagnosed patients with NSGCTs present with clinical stage II disease. Patients with stage IIA disease may be treated with primary RPLND. Alternatively, these patients may be treated with cisplatin-based chemotherapy. Patients with stage IIB and IIC disease are best initially managed with cisplatin-based chemotherapy.

Stage III Patients who present with stage III GCT (seminoma or NSGCT) are treated with cisplatin-based chemotherapy. These patients are classified into good-, intermediate-, or poor-risk categories using the International Germ Cell Consensus Classification system, which is based on clinical factors including histology, site of primary, the presence of nonpulmonary visceral metastatic disease, and the level of postorchietomy serum tumor markers (Table 88-1). Most patients with stage III GCT present with good-risk disease; >90% will be cured. The remainder present with intermediate-risk or poor-risk disease, associated with 5-year survival rates of ~80% and 50%, respectively. Select patients with rapidly progressive metastatic disease and life-threatening symptoms such as hemoptysis in whom there is a high clinical suspicion of GCT should emergently initiate cisplatin-based chemotherapy, even without a tissue diagnosis.

Chemotherapy The development of cisplatin-based chemotherapy represents an important advance in cancer medicine. Through a series of carefully performed clinical trials with the aim of maximizing cure while minimizing the extent of treatment, the chemotherapy approach to the treatment of these patients has been standardized. Patients with good-risk metastatic GCT are treated with either three cycles of BEP or four cycles of etoposide and cisplatin (EP). Patients with intermediate- and poor-risk metastatic disease are treated with either four cycles of BEP or four cycles of etoposide, ifosfamide, and cisplatin (VIP). Maintaining dose and schedule is important, as dose modifications and delays have been associated with inferior outcomes. Serum tumor markers should be monitored throughout treatment and should normalize during or after treatment. Cisplatin-based chemotherapy is associated with myelosuppression, nausea and vomiting, and alopecia. Cisplatin may result in nephrotoxicity, ototoxicity, and peripheral neuropathy. Bleomycin may result in pulmonary toxicity, and risk factors for this include age >40 , renal failure, tobacco use, and the cumulative dose of bleomycin received. For patients at increased risk of bleomycin-induced pneumonitis, non-bleomycin-containing regimens as noted above may be given. Cisplatin-based chemotherapy is also associated with sterility. Approximately 30% of newly diagnosed testicular GCT patients have severe oligospermia or azoospermia. For the remainder with normal baseline spermatogenesis who receive cisplatin-based chemotherapy, all will be azoospermic at the completion of therapy. Approximately 80% of these patients will recover spermatogenesis over a period of several years. For this reason, prechemotherapy sperm banking should be offered to all patients treated with chemotherapy.

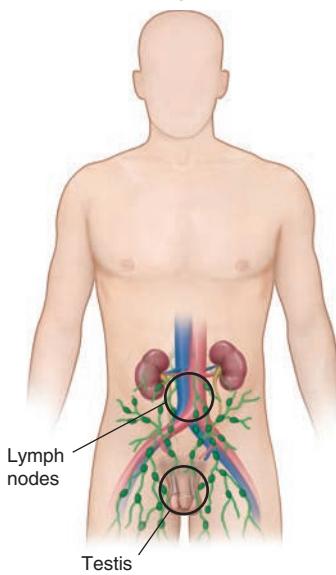
Stage 1



	Seminoma	NSGCT
Stage IA Testis only, no lymphovascular invasion	Active surveillance; or, Adjuvant carboplatin x 1 or 2 cycles; or, Adjuvant para-aortic RT	Active surveillance; or, Nerve-sparing RPLND; or Adjuvant BEP x 1 cycle
Stage IB Testis only, with lymphovascular invasion or invasion of spermatic cord or scrotum	Active surveillance; or, Adjuvant carboplatin x 1 or 2 cycles; or, Adjuvant para-aortic RT	Active surveillance; or, Adjuvant BEP x 1 cycle; or Nerve-sparing RPLND
Stage IS Elevated serum tumor markers post-orchiectomy	BEP x 3 cycles; or, EP x 4 cycles	BEP x 3 cycles; or, EP x 4 cycles

A

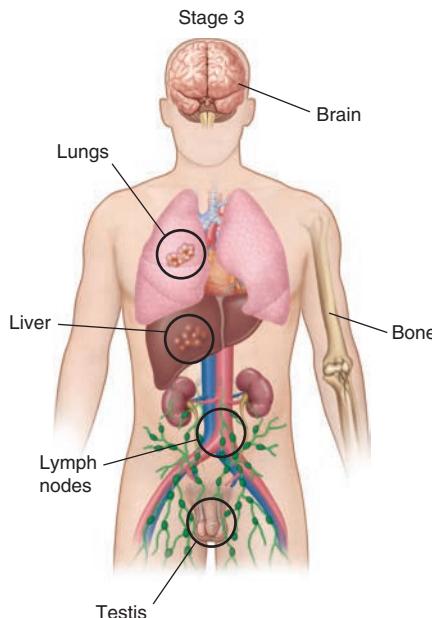
Stage 2



	Seminoma	NSGCT
Stage IIA N1: nodes ≤ 2 cm	Para-aortic and ipsilateral iliac RT; or, BEP x 3 cycles or EP x 4 cycles	Nerve-sparing RPLND; or, BEP x 3 cycles or EP x 4 cycles
Stage IIB N2: nodes > 2 to 5 cm	BEP x 3 cycles or EP x 4 cycles; or, Para-aortic and ipsilateral iliac RT	BEP x 3 cycles or EP x 4 cycles +/- postchemotherapy RPLND
Stage IIC N3: nodes > 5 cm	BEP x 3 cycles or EP x 4 cycles	BEP x 3 cycles or EP x 4 cycles +/- postchemotherapy RPLND

B

FIGURE 88-1 Stage-based management of testicular germ cell tumor.



	Seminoma	NSGCT
Stage IIIA (good-risk)	BEP x 3 cycles; or, EP x 4 cycles	BEP x 3 cycles; or, EP x 4 cycles; +/- Postchemotherapy surgery
Stage IIIB (intermediate-risk)	BEP x 4 cycles; or, VIP x 4 cycles	BEP x 4 cycles; or, VIP x 4 cycles +/- Postchemotherapy surgery
Stage IIIC (poor-risk)	N/A	BEP x 4 cycles; or, VIP x 4 cycles +/- Postchemotherapy surgery

Abbreviations: BEP, bleomycin, etoposide, cisplatin; EP, etoposide, cisplatin; N/A, not applicable; NSGCT, nonseminomatous germ cell tumor; RPLND, retroperitoneal lymph node dissection; RT, radiation therapy; VIP, etoposide, ifosfamide, cisplatin.

C

FIGURE 88-1 (Continued)

TABLE 88-1 International Germ Cell Consensus Classification System

RISK GROUP	SEMINOMA	NSGCT
Good	Any primary site; and normal AFP, any hCG, any LDH; and nonpulmonary visceral metastases absent	Gonadal or retroperitoneal primary; and nonpulmonary visceral metastases absent; and AFP <1000 ng/mL; and hCG <5000 mIU/mL; and LDH <1.5 × ULN
Intermediate	Any primary site; and normal AFP, any hCG, any LDH; and nonpulmonary visceral metastases present	Gonadal or retroperitoneal primary; and nonpulmonary visceral metastases absent; and one of the following: AFP 1000–10,000 ng/mL hCG 5000–50,000 mIU/mL LDH 1.5–10 × ULN
Poor	N/A	Mediastinal primary; or nonpulmonary visceral metastases present; or one of the following: AFP >10,000 ng/mL hCG >50,000 mIU/mL LDH >10 × ULN

Abbreviations: AFP, α fetoprotein; hCG, human chorionic gonadotropin; LDH, lactate dehydrogenase; N/A, not applicable; NSGCT, nonseminomatous germ cell tumor; ULN, upper limit normal. Nonpulmonary visceral metastases include liver, bone, and brain.

Source: Reproduced with permission from International Germ Cell Cancer Collaborative Group: International Germ-Cell Consensus Classification: A prognostic factor based staging system for metastatic germ cell tumors. J Clin Oncol 15:594, 1997.

Postchemotherapy Surgery Upon completion of cisplatin-based chemotherapy, many patients with normalized serum tumor markers will have radiographic evidence of residual masses. In approximately half of patients with NSGCT, the residual mass is composed of necrosis and/or fibrosis. About 40% will have residual teratoma and only 10% will have residual viable nonteratomatous GCT. Unfortunately, radiographic imaging cannot accurately differentiate between these entities. For this reason, NSGCT patients with residual masses after chemotherapy undergo resection of all sites of disease. This most commonly includes a postchemotherapy RPLND. However, thoracotomy and neck dissection are required in some patients. If the patients are found to have residual necrosis or teratoma, no additional therapy is required. However, for patients with residual viable nonteratomatous GCT, two additional cycles of chemotherapy are frequently administered. It should be noted that in most centers, patients with minimal residual tumors defined as retroperitoneal lymph nodes of ≤ 1 cm will forego postchemotherapy RPLND. Patients who experience normalization of serum tumor markers with first-line chemotherapy but have enlarging tumors, most often cystic masses in the retroperitoneum, may have “growing teratoma syndrome.” These patients are best approached with surgery.

For patients with metastatic seminoma, most residual masses are necrotic and do not harbor viable tumor. Patients with residual masses of 3 cm or less may be observed without surgery. For patients with residual masses >3 cm, fluorodeoxyglucose (FDG)-PET may be used to distinguish necrosis from viable seminoma and identify patients who should be considered for postchemotherapy surgery or short interval imaging.

■ RELAPSED DISEASE

Approximately 20–30% of patients with metastatic GCTs treated with cisplatin-based chemotherapy will not achieve durable disease control. Most of these patients will experience disease progression within 2 years following completion of chemotherapy. The International Prognostic Factors Study Group developed a risk stratification classification system for patients in first relapse. Contributors to a worsened prognosis include NSGCT histology, extragonadal primary, incomplete response to first-line chemotherapy, time to relapse of 3 months or less, level of serum tumor markers at relapse, and the presence of nonpulmonary visceral metastatic disease.

Patients in first relapse may be treated with either conventional-dose salvage chemotherapy or high-dose salvage chemotherapy with autologous stem cell rescue. There is controversy concerning which approach is optimal. Some institutions advocate for risk stratification, with more favorable prognosis patients receiving conventional-dose chemotherapy and worse prognosis patients receiving high-dose chemotherapy. The most commonly utilized conventional-dose regimen includes paclitaxel, ifosfamide, and cisplatin (TIP). In one study of TIP in patients with more favorable-risk disease, approximately two-thirds experienced 2-year progression-free survival. High-dose chemotherapy consists of initial salvage therapy followed by stem cell harvest and then two or three cycles of high-dose carboplatin and etoposide (CE) with stem cell rescue. The largest series of patients treated with high-dose chemotherapy was reported by researchers at Indiana University where this approach is considered standard for most patients in first relapse regardless of risk classification. In their study, ~70% of patients in first relapse achieved durable progression-free survival. A large retrospective analysis has compared conventional-dose salvage chemotherapy to high-dose salvage chemotherapy in patients in first relapse. This study reports a more favorable outcome with high-dose salvage chemotherapy across nearly all risk groups. However, given the retrospective nature of this study and the controversy concerning optimal approaches, an international randomized trial comparing conventional-dose chemotherapy (TIP) to high-dose chemotherapy with autologous stem cell rescue (TI-CE) is underway.

Some patients who experience disease progression after conventional-dose salvage chemotherapy may successfully be treated with high-dose salvage chemotherapy with autologous stem cell rescue. Patients with disease progression after high-dose salvage chemotherapy may be treated with subsequent chemotherapy regimens that include gemcitabine/oxaliplatin, gemcitabine/paclitaxel, epirubicin/cisplatin, and oral etoposide. While these patients may benefit from third-line chemotherapy, few will achieve durable disease control. Select patients with relapsed but resectable disease may be candidates for salvage or so-called “desperation” surgery. Studies of molecularly targeted agents and immune checkpoint inhibitors in this population have to date been generally disappointing.

Patients who experience disease progression >2 years after chemotherapy are considered to have “late relapse.” Late relapse appears to have a different biology than early relapse. These patients tend to have more chemotherapy-resistant disease. Patients with late relapse usually have NSGCT with elevation of serum AFP. Many of these patients experience recurrence in the retroperitoneum many years after first-line chemotherapy, and this likely represents residual retroperitoneal disease that was not controlled after first-line therapy. These patients are best approached with salvage surgery.

■ EXTRAGONADAL GCTS

Approximately 5% of patients who present with GCTs have extragonadal primaries. These mainly originate in the mediastinum or retroperitoneum. Patients suspected of extragonadal GCT should undergo scrotal ultrasound to exclude a gonadal primary. Extranodal seminomas have a similar excellent prognosis as their gonadal counterparts and are approached the same. Mediastinal NSGCTs are classified as poor risk and are treated with either four cycles of BEP or four cycles of VIP. These patients frequently require postchemotherapy thoracic surgery for residual disease. For this reason, some advocate avoiding bleomycin in this patient population. Klinefelter’s syndrome is associated with an increased risk of mediastinal NSGCTs. Rarely, mediastinal

NSGCTs are associated with hematologic disorders including acute myelogenous leukemia. NSGCTs arising in the retroperitoneum do not have a worse prognosis than their gonadal counterparts. Many patients who present with extragonadal GCTs will undergo core needle biopsy for diagnosis. However, select patients with extragonadal tumors and definitive elevation of serum tumor markers may initiate chemotherapy without a tissue diagnosis.

Cancers of unknown primary are defined as histologically proven metastatic malignancy in which the primary site is not obvious. A subgroup of patients with cancer of unknown primary have occult GCTs. Male gender, age <65 years, midline tumors, and nonsmoking status increase the likelihood of this presentation. Pathology may demonstrate a poorly differentiated malignant neoplasm. Immunohistochemical staining is used to exclude lymphoma. Tumor may be analyzed by fluorescence in situ hybridization for i(12p), which confirms the diagnosis. Even if the diagnosis is not certain, patients should be treated with cisplatin-based chemotherapy, which will cure up to 20% of this patient group.

■ TESTICULAR NON-GERM CELL TUMORS

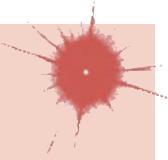
Rarely, patients may develop testicular non-GCTs. These include non-Hodgkin’s lymphoma, most commonly occurring in men over the age of 50; sex cord stromal tumors including Leydig cell tumors and Sertoli cell tumors; mesothelioma of the tunica vaginalis; and paratesticular sarcoma. Metastasis to the testis is rare, most commonly occurring in patients with advanced prostate cancer and melanoma.

■ SURVIVORSHIP AND LATE EFFECTS

Because most patients with testicular GCT will experience long-term survival, survivorship care is important. Since many of these patients will be followed by primary care physicians, an understanding of the physical, psychological, and social late effects is important. Late effects are defined as health problems that occur months or years after a disease is diagnosed or after treatment has ended. Late effects may be related to the underlying cancer or to the treatment the patient received. In long-term survivors of testicular GCT, increased cardiovascular risk and increased secondary malignancies have been reported. Patients treated with cisplatin-based chemotherapy have an increased risk of hypertension, hyperlipidemia, metabolic syndrome, and cardiovascular events. Patients treated with high cumulative doses of etoposide (e.g., patients who receive standard chemotherapy, relapse, and then receive salvage high-dose chemotherapy) may experience up to a 1–2% risk of developing acute myelogenous leukemia, typically 2–3 years after completing therapy and associated with an 11q23 translocation. Patients treated with radiation therapy, cisplatin-based chemotherapy, or both have an increased risk of developing secondary solid malignancies.

■ FURTHER READING

- EINHORN LH et al: High-dose chemotherapy and stem-cell rescue for metastatic germ cell tumors. *N Engl J Med* 357:340, 2007.
- FELDMAN DR et al: Medical treatment of advanced testicular cancer. *JAMA* 299:272, 2008.
- FUNG C et al: Testicular cancer survivorship. *J Natl Compr Canc Netw* 17:1557, 2019.
- HANNA NH, EINHORN LH: Testicular cancer—discoveries and updates. *N Engl J Med* 371:2005, 2014.
- INTERNATIONAL PROGNOSTIC FACTORS STUDY GROUP et al: Prognostic factors in patients with metastatic germ cell tumors who experienced treatment failure with cisplatin-based first-line chemotherapy. *J Clin Oncol* 28:4906, 2010.
- KOLLMANSBERGER C et al: Patterns of relapse in patients with clinical stage 1 testicular cancer managed with active surveillance. *J Clin Oncol* 33:51, 2015.
- LORCH A et al: Conventional-dose versus high-dose chemotherapy as first salvage treatment in male patients with metastatic germ cell tumors: Evidence from a large international database. *J Clin Oncol* 29:2178, 2011.
- PLUTA J et al: Identification of 22 novel susceptibility loci associated with testicular germ cell tumors. *Nat Commun* 12:4487, 2021.



OVARIAN CANCER

INCIDENCE AND PATHOLOGY

Ovarian cancer remains a leading cause of cancer deaths in American women, ranking behind lung, breast, colon, and pancreatic cancers. The ovary is responsible for hormone and egg production. Between menarche (11–13 years) and menopause (45–55 years), the ovary is responsible for follicle maturation associated with egg maturation, ovulation, and cyclical sex steroid hormone production. These complex biologic functions are linked to stromal and germ cells within the ovary. Cells of the ovary can be broadly grouped into stromal cells and ovarian germ cells and the enveloping epithelial cells. Malignancies arising in each group include multiple histologic variants, each with unique neoplastic behaviors. Epithelial tumors are the most common histologic variant of ovarian neoplasms; they may be benign (50%), frankly malignant (33%), or of borderline malignancy of low malignant potential (16%). In adnexal masses detected by imaging or physical exam, age influences risk of malignancy; tumors in younger women are more likely benign. In the malignant group, the most common tumors are epithelial. In the group of the ovarian epithelial malignancies are the serous tumors (60–70%), mucinous tumors (10%), endometrioid tumors (10–15%), and clear cell tumors (10–15%) tumors. The distribution of histologic types varies in different parts of the world. The less common stromal tumors arise from the ancillary, supportive cells such as steroid hormone-producing cells and likewise have different phenotypes and clinical presentations. Most stromal tumors do not produce estrogen, but ectopic hormone production can be seen in certain subtypes. Tumors arising in the ovarian germ cell lineage are generally similar in biology and behavior to testicular tumors in males, although their intraperitoneal location alters some metastatic behaviors (Chap. 88). Ovarian tissue may also host metastatic tumors arising from breast, colon, gastric, and pancreatic primaries. Bilateral ovarian masses from metastatic mucin-secreting gastrointestinal cancers are termed *Krukenberg tumors*. A survey of other potential primaries is commonly required during the diagnostic workup of ovarian masses.

OVARIAN CANCER OF EPITHELIAL ORIGIN

Epidemiology An American woman has approximately a 1 in 72 lifetime risk (1.6%) of developing ovarian cancer, with the majority of affected women developing epithelial tumors. In 2021 in the United States, ~21,500 cases of ovarian cancer are expected to be diagnosed, with >14,000 deaths. Sporadic (not familial) epithelial tumors of the ovary have a peak incidence in women in their fifties and sixties, although age at presentation ranges from the third decade to the eighties and nineties. Ovarian cancer risk has been linked to an interactive mixture of epidemiologic, environmental, and genetic factors. Nulliparity, obesity, diet, infertility treatments, and possibly hormone replacement therapy have all been linked to an increase in risk. Protective factors include the use of oral contraceptives, multiparity, tubal ligation, aspirin use, and breast-feeding. Other epidemiologic factors such as the historical use of perineal talc agents remain controversial. The mechanisms underlying the various protective factors are largely unknown, but theories include suppression of ovulation, modulation of gonadotropins and progestins, and perhaps reduction of ovarian inflammation and damage associated with the repair of the ovarian cortex associated with ovulation.

Genetics and Pathogenesis Ovarian cancers are divided into type 1 cancers and the more aggressive type 2 variant. The type 1 cancers are characterized by low-grade histology and generally indolent behavior. These tumors include the low malignant potential tumors,

low-grade endometrial and mucinous histologies, and clear cell cancers (which are more aggressive). Genetic alterations in type 1 cancers include mutations in *KRAS*, *BRAF*, *PTEN*, and *PIK3CA*. In contrast, studies have implicated serial genetic changes in the fallopian tube as the actual site of origin for most type 2, high-grade serous epithelial ovarian cancers. These aggressive tumors are more common and linked to losses in *TP53* and defective DNA repair. Carcinoma *in situ* has been identified in the tubal epithelium with early losses in *TP53* and the *BRCA1/BRCA2* gene function characterizing early tubal intraepithelial cancers. Following these early genetic events, additional mutations in these transformed cells lead to tumor cell shedding, metastasis, and invasion. These type 2, poorly differentiated, serous cancer cells can then spread to the ovaries and the peritoneal cavity, aided by the ovarian cancer cell's affinity for mesothelin-expressing cells.

Type 2 serous ovarian cancer is classically a disease characterized by widespread amplifications and deletions rather than single-gene point mutations or common gene fusions. In the Tumor Genome Atlas, loss of tumor-suppressor gene *TP53* function is present in >95% of serous ovarian cancers. Damage to homologous DNA repair genes, especially *BRCA1* and *BRCA2*, is also common in these tumors. Low prevalence but statistically recurrent somatic mutations in seven other genes including *NF1*, *RB1*, and *CDK12* were also seen. The most common heritable abnormality linked to ovarian cancer is a germline mutation in either *BRCA1* (chromosome 17q12–21) or *BRCA2* (chromosome 13q12–13). These genes are essential parts of the homologous DNA repair machinery for double-stranded DNA break repair. Individuals inheriting a single copy of a mutant allele have an increased lifetime risk of breast (46–87% for *BRCA1*; 38–84% for *BRCA2*) and ovarian cancer (39–63% for *BRCA1*; 16.5–27% for *BRCA2*). Many of these women have a family history that includes multiple cases of breast and/or ovarian cancer of at an early age. Male breast cancer, pancreatic cancer, and prostate cancer are also linked to familial *BRCA2* mutations. The most common malignancy in women carrying germline *BRCA1/2* mutations is breast carcinoma, although women harboring germline *BRCA1* mutations also have a marked increased risk of developing ovarian malignancies in their forties and fifties. Women harboring a mutation in *BRCA2* have a lower penetrance of ovarian cancer with onset typically in their fifties or sixties. Other uncommon germline mutation of other genes encoding proteins linked to homologous DNA repair (e.g., *PALB2*) can also contribute to cancer risk, although the frequency of mutation and magnitude of risk increment are much lower and not well defined. Screening studies, even in the mutated *BRCA1/2* families, suggest that any of the available screening techniques, including structured, serial evaluation of the CA-125 serum marker and transvaginal ultrasound, remain insufficient to reliably detect early-stage ovarian cancer in prospective testing. Germline *BRCA1/2* testing is recommended for all incident epithelial ovarian cancers to detect probands for therapeutic intervention and identify relatives at risk. Women with these high-risk germline mutations are advised to undergo prophylactic removal of fallopian tubes and ovaries after completing childbearing, ideally before age 40. Early prophylactic salpingo-oophorectomy is highly protective. Salpingo-oophorectomy also appears to protect these women from subsequent breast cancer (risk reduction 50%). Prophylactic salpingectomy is almost certainly a key part of any surgical prophylaxis strategy for ovarian cancer prevention, but the benefits of isolated oophorectomy on either ovarian or breast cancer risk have not yet been clearly defined. Although less common, ovarian cancer is also another familial form of cancer (along with colorectal and endometrial cancer) that may develop in women with type II Lynch syndrome caused by mutations in one of the DNA mismatch repair genes (*MSH2*, *MLH1*, *MLH6*, *PMS1*, *PMS2*). Ovarian cancer may appear in women younger than 50 years of age in this syndrome.

Neoplasms of the ovary tend to be painless unless they undergo torsion. Nonspecific gastrointestinal symptoms like bloating and early satiety are common at presentation, probably related to compression of local organs or due to symptoms from metastatic disease. Women with ovarian tumors also may have an increased incidence of symptoms including pelvic discomfort, bloating, and perhaps changes in urinary or bowel pattern. Unfortunately, all of these symptoms are common in

TABLE 89-1 Staging and Survival in Gynecologic Malignancies

STAGE	OVARIAN	5-YEAR SURVIVAL, %	ENDOMETRIAL	5-YEAR SURVIVAL, %	CERVIX	5-YEAR SURVIVAL, %
0	—		—		Carcinoma in situ	100
I	Confined to ovary	88–95	Confined to corpus	>90	Confined to uterus	85
II	Confined to pelvic organs	70–80	Involves corpus and cervix	~75	Invades beyond uterus but not to pelvic wall	65
III	Intra-abdominal spread to omentum, diaphragm, or lymph nodes	20–40	Extends outside the uterus but not outside the true pelvis	45–60	Extends to pelvic wall and/or lower third of vagina, or hydronephrosis	35
IV	Spread outside abdominal cavity, parenchymal spread, and pleural effusion cytology	17	Extends outside the true pelvis or involves the bladder or rectum	~20	Invades mucosa of bladder or rectum or extends beyond the true pelvis	7

primary care and are frequently dismissed by either the woman or her health care team until later stages of disease. The pathogenic factors and timing of spread beyond the ovary are still not well understood. The most common symptoms at presentation of advanced disease include a period of progressive complaints of nausea, early satiety, bloating, indigestion, constipation, and abdominal pain. Signs include the rapid increase in abdominal girth due to the accumulation of ascites that typically alerts the patient and her physician that the concurrent gastrointestinal symptoms are likely associated with malignant pathology. Radiologic evaluation typically demonstrates a complex adnexal mass with ascites, carcinomatosis, and pelvic, para-aortic and mesenteric adenopathy in advanced disease. Positron emission tomography (PET) scans are generally not required. Laboratory evaluation often demonstrates a markedly elevated CA-125, a shed mucin component (MUC16) associated with, but not specific for, ovarian cancer. Ovarian cancers are divided into four stages, with stage I tumors confined to the ovary, stage II malignancies confined to the pelvis, and stage III confined to the peritoneal cavity and retroperitoneal nodes (**Table 89-1**). These three stages are subdivided, with the most common presentation, stage IIIC, defined as tumors with bulky intraperitoneal disease or positive lymph node involvement. About 70% of women present with stage III disease. Stage IV disease includes women with parenchymal metastases (liver, lung, spleen) or, alternatively, abdominal wall or pleural disease. The 30% of patients not presenting with stage III disease are roughly evenly distributed among the other stages.

Screening Ovarian cancer is a highly lethal condition. It is curable in early stages but seldom curable in advanced stages; hence, screening continues to be of considerable interest. Early-stage tumors often secrete excessive amounts of normal proteins that can be measured in the serum such as CA-125, mesothelin, and HE-4. Nevertheless, the incidence of ovarian cancer in the middle-aged female population is very low, with only ~1 in 2000 women between the ages of 50 and 60 carrying an asymptomatic and undetected tumor. Thus, effective screening techniques must be both sensitive and highly specific to minimize the number of false positives. Panels of serum markers have not improved on CA-125 alone, nor have risk assessment strategies using algorithms with multiple CA-125 measurements over time. No other screening strategies have been any more successful to date. Some large studies have suggested that low-specificity screening might even worsen mortality in the screened population. Screening for ovarian cancer is currently not recommended outside of a clinical trial, but large ongoing clinical trials are studying algorithmic detection by serial sampling strategies.

TREATMENT

Ovarian Cancer

Epithelial ovarian cancer can be divided into distinct “disease states” for the purpose of treatment selection, as shown in **Fig. 89-1**. Surgery by a skilled gynecologic oncologist remains the preferred initial therapy for ovarian cancer. However, the amount of residual visible cancer at the end of a primary operation is

strongly predictive of outcome and is paired with histology, grade, and stage to determine prognosis and treatment. In women presenting with a localized ovarian mass, the principal diagnostic and therapeutic maneuver is abdominal surgery to determine if the tumor is benign or malignant. In the event that the tumor is malignant, the surgical specimen will determine if the tumor arises in the ovary or is a site of metastatic disease. Metastatic disease to the ovary can be seen from primary tumors of the colon, appendix, stomach (Krukenberg tumors), and breast. Needle biopsy is contraindicated to avoid malignant contamination of the peritoneal cavity with malignant cells. Typically, women undergo laparoscopic evaluation and unilateral salpingo-oophorectomy for diagnostic purposes. If pathology reveals a primary ovarian malignancy or the laparoscopy proves disseminated disease is present, then the procedure should be followed by a total hysterectomy, removal of the remaining tube and ovary, omentectomy, and pelvic node sampling along with biopsies of the peritoneal cavity and diaphragms. This extensive surgical procedure is performed because ~30% of tumors that, by visual inspection, appear to be confined to the ovary have already disseminated to the peritoneal cavity and/or surrounding lymph nodes. As with axillary dissections in breast cancer, node sampling is diagnostic, but full lymphadenectomy appears to provide little or no additional therapeutic advantage over nodal sampling. The target outcome of an ovarian cancer surgery is always an R0 resection, with no visible residual cancer. The less favorable “optimal resection” (no disease >1 cm in size) is still clinically useful, and the prognosis of those patients is much better than that of patients who are left with >1 cm of disease at the end of surgery. These “suboptimally debulked” patients derive very little benefit from their surgery. If a suboptimal debulking is anticipated, the surgery should be delayed until after several cycles of neoadjuvant chemotherapy. Such “interval debulking” surgery achieves similar results to primary surgery with diminished surgical morbidity and more timely chemotherapy. Patients without gross residual disease after resection have a median survival in excess of 60 months, compared to 28–42 months for those left with macroscopic tumor or those undergoing interval debulking, regardless of treatment strategy.

After appropriate surgical treatment, primary chemotherapy will consist of combination treatment with paclitaxel and carboplatin. Primary chemotherapy can be delivered intravenously, or alternatively, some therapy can be directly administered into the peritoneal cavity via an indwelling catheter. Some, but not all, randomized studies have demonstrated improved survival with intraperitoneal (IP) therapy. The IP approach is technically more difficult and is increasingly replaced by carboplatin and paclitaxel, which appears to offer similar results.

With optimal debulking surgery and platinum-based chemotherapy (usually carboplatin dosed to an area under the curve [AUC] of 6.0 plus paclitaxel 175 mg/m² by 3-h infusion in monthly cycles), 70% of women who present with advanced-stage tumors show tumor reduction, and 40–50% experience a complete remission with normalization of their CA-125, CT scans, and physical

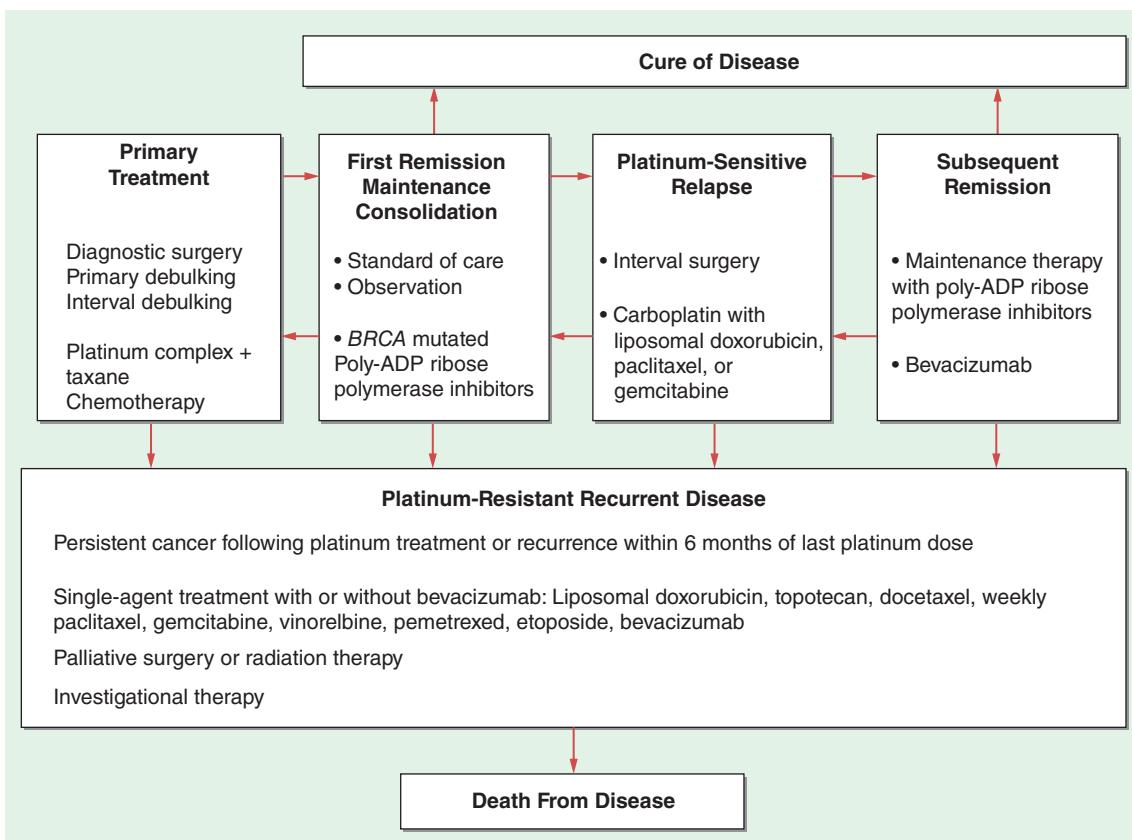


FIGURE 89-1 Disease states model of epithelial ovarian cancer and its treatment. Each box represents a relatively homogenous group of patients who share a palette of potential treatment choices and have a similar prognosis. The arrows indicate that a single patient may move from one state to another during the course of her illness, and the choice of treatments will become different in her new disease state.

examination. Poly-ADP ribose polymerase inhibitors (PARPi) such as niraparib or olaparib, when administered at the completion of intravenous chemotherapy, appear to substantially delay recurrence and probably provide survival advantages as well. In the majority of patients, disease still recurs within 1–4 years from the completion of their primary therapy. CA-125 levels often increase as a first sign of relapse, and CT scan findings are confirmatory. Recurrent disease is often successfully managed for years, but rarely cured, with a variety of chemotherapeutic agents. Additional surgical therapy does not appear to extend survival in randomized trials. Patients with a treatment-free interval are often best treated with additional platinum doublets, combining carboplatin with liposomal doxorubicin, gemcitabine, or a taxane. Eventually all women who experience relapse develop chemotherapy-refractory disease. Refractory ascites, poor bowel motility, and obstruction or tumor-infiltrated aperistaltic bowel are all common premorbid events. Limited surgery to relieve intestinal obstruction, localized radiation therapy to relieve pressure or pain from masses, or palliative chemotherapy may be helpful. Agents with >15% response rates include gemcitabine, topotecan, liposomal doxorubicin, and bevacizumab. Five-year survival correlates with the stage of disease: stage I, 90–95%; stage II, 70–80%; stage III, 25–40%; stage IV, 10–15% (Table 89-1). Prognosis is also influenced by histologic grade: 5-year survival is 88% for well-differentiated tumors, 58% for moderately differentiated tumors, and 27% for poorly differentiated tumors. Histologic type has less influence on outcome.

their tumors (10-year survival may approach 98%), although recurrence is not uncommon. Certain features, such as micropapillary histology and microinvasion, are linked to more aggressive behavior. Tumors of low malignant potential should be carefully distinguished from grade 1 serous carcinomas. Borderline tumor patients are managed primarily by surgery; chemotherapy and radiation therapy do not substantially alter survival.

Stromal Tumors Approximately 7% of ovarian neoplasms are stromal tumors, with ~1800 cases expected each year in the United States. Ovarian stromal tumors or sex cord tumors are most common in women in their fifties or sixties, but tumors can present at any age. These tumors arise from the mesenchymal components of the ovary, including both steroid-producing cells and fibroblasts. Most of these tumors are indolent tumors with limited metastatic potential and present as unilateral solid masses. These tumors primarily are discovered by the detection of an abdominal mass, sometimes with abdominal pain due to ovarian torsion, intratumoral hemorrhage, or rupture. Rarely, stromal tumors can produce estrogen and present with breast tenderness as well as precocious puberty in children, menstrual disturbances in reproductively active women, or postmenopausal bleeding. In some women, estrogen-associated secondary malignancies, such as endometrial or breast cancer, may present as synchronous malignancies. Sertoli-Leydig tumors often present with hirsutism and virilization due to increased production of androgens. Hormonally inert tumors include fibromas, which present as solitary masses often in association with ascites and occasionally hydrothorax, also known as Meig's syndrome. A subset of these tumors present in individuals with a variety of inherited disorders that predispose them to mesenchymal neoplasia including Ollier's disease (juvenile granulosa cell tumors) and Peutz-Jeghers syndrome (ovarian sex cord tumors). The treatment of these tumors is almost exclusively by surgical resection, without adjuvant chemotherapy. Chemotherapy with carboplatin and

■ UNCOMMON OVARIAN TUMORS

Low Malignant Potential Tumors (Borderline Tumors) These type 1 tumors are found in younger women (age 30–50 years) and indolent in behavior, and few of these patients will succumb to

paclitaxel is generally reserved for either unresectable or multiply recurrent tumors.

Germ Cell Tumors of the Ovary Germ cell tumors, like their counterparts in the testis, are cancers of germ cells. These totipotent cells contain the programming for differentiation to essentially all tissue types, and hence, the germ cell tumors include a histologic menagerie of bizarre tumors, including benign teratomas (dermoid cysts) and a variety of malignant tumors, such as dysgerminoma, immature teratomas, yolk sac malignancies, and choriocarcinomas. Benign teratoma (or dermoid cyst) is the most common germ cell neoplasm of the ovary and often presents in young women. These tumors include a complex mixture of differentiated tissue including tissues from all three germ layers. In older women, these differentiated tumors can develop malignant transformation, most commonly squamous cell carcinomas. Malignant germ cell tumors include dysgerminomas, yolk sac tumors, immature teratomas, and embryonal and choriocarcinomas. Germ cell tumors can present at all ages, but the peak age of presentation tends to be in adolescents. Typically, these tumors will become large ovarian masses, which eventually present as palpable low abdominal or pelvic masses. Like sex cord tumors, torsion or hemorrhage may present urgently or emergently as acute abdominal pain. Some germ cell tumors produce elevated levels of human chorionic gonadotropin (hCG) or α -fetoprotein (AFP). Unlike epithelial ovarian cancer, these tumors have a higher proclivity for nodal or hematogenous metastases. Germ cell tumors typically present in women who are of childbearing age, and because bilateral tumors are uncommon (except in dysgerminoma, 10–15%), the typical treatment is unilateral oophorectomy or salpingo-oophorectomy with lymph node sampling. Most commonly, women with advanced malignant germ cell tumors typically receive bleomycin, etoposide, and cisplatin (BEP) chemotherapy, in an analogous fashion to the treatment of testicular cancers. In the majority of these women, even those with advanced-stage disease, cure is expected. Dysgerminoma is the ovarian counterpart of testicular seminoma and is highly curable. Although the tumor is highly radiation-sensitive, radiation produces infertility in many patients. BEP chemotherapy is as effective or more so without causing infertility.

FALLOPIAN TUBE CANCER

Transport of the egg to the uterus occurs through the fallopian tube, with the distal ends of these tubes composed of fimbriae that drape about the ovarian surface and capture the egg as it erupts from the ovarian cortex. As described above, the majority of type 2 ovarian cancers are now thought to arise from the tubal epithelium. As might be expected, fallopian tube malignancies are typically of serous histology and share the same biology and recommended treatment as serous ovarian cancer. These tumors often present as clinically isolated adnexal masses, but like ovarian cancer, these tumors spread relatively early throughout the peritoneal cavity. Fallopian tubal cancers have a natural history and treatment that are essentially identical to ovarian cancer (Table 89-1).

CERVICAL CANCER

ETIOLOGY AND GENETICS

Cervical cancer is the second most common and the most lethal malignancy in women worldwide. Infection with high-risk strains of human papillomavirus (HPV) is the primary neoplastic-initiating event in the vast majority of women with invasive cervical cancer. This double-stranded DNA virus infects epithelium near the transformation zone of the cervix where underlying columnar epithelium becomes squamous epithelium. More than 60 types of HPV are known, with ~20 types having the ability to generate high-grade dysplasia and malignancy. HPV16 and 18 are the types most frequently associated with high-grade dysplasia, but types 31, 33, 35, 52, and 58 are also considered to be high-risk variants. The large majority of sexually active adults are exposed to HPV, and most women clear the infection without specific intervention. The 8-kb HPV genome encodes seven

early genes, most notably *E6* and *E7*, which can bind to *RB* and *p53*, respectively. High-risk types of HPV encode *E6* and *E7* molecules that are particularly effective at inhibiting the normal cell cycle checkpoint functions of these regulatory proteins, leading to immortalization but not full transformation of cervical epithelium. A minority of women will fail to clear the infection, with subsequent HPV integration into the host genome. Over as little as a few months to several years, some of these persistently infected women develop worsening dysplasia, a premalignant condition that, untreated, can progress to cervical carcinoma. Complete transformation to cancer occurs over a period of years and almost certainly requires the acquisition of other poorly defined genetic mutations within the infected and immortalized epithelium.

In 2018, ~570,000 new cases of cervical cancer occurred worldwide, with an estimated 311,000 deaths. Cancer incidence is particularly high in women residing in Central and South America, the Caribbean, and southern and eastern Africa. The mortality rate is disproportionately high in Africa. In the United States, an estimated 14,480 women will be diagnosed with cervical cancer in 2021 and ~4290 women will die of the disease.

In the integrated genomic characterization of cervical cancer by The Cancer Genome Atlas (TCGA), integration of HPV sequences was found in all of the HPV18-linked cancers and over three-quarters of the HPV16 cancers. The cervical tumors also showed a characteristic APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; a family of cytidine deaminases that edit DNA and are endogenous mutagenic enzymes) pattern of mutagenesis, with *ERBB3*, *CASP8*, and *TGFRB2* identified as significantly mutated genes presumably linked to progression from dysplasia to carcinoma. In the much smaller number of HPV-negative cancers, which are more common in older women, mutations in oncogenes *KRAS*, *ARID1A*, and *PTEN* were frequently seen. The clinical behavior of these cancers is likely to be different.

HPV INFECTION AND PREVENTION

The Pap smear is the primary detection method for asymptomatic preinvasive cervical dysplasia of squamous epithelial lining during a gynecologic exam. Because the delay between dysplasia and frank cervical cancer is years long, annual (or longer) screening and prevention strategies that detect precancerous dysplasia and carcinoma *in situ* can be implemented successfully. Annual or biannual cervical scraping for cytology (Pap smear) is highly effective in reducing the incidence of cervical cancer by early detection and subsequent surgical treatment of premalignant disease. The incorporation of HPV testing by polymerase chain reaction (PCR) or other molecular techniques increases the sensitivity of detecting cervical pathology but at the cost of lower sensitivity in that it identifies many women with transient infections who require no specific medical intervention. Unfortunately, both the collection of a Pap smear and its cytologic evaluation require infrastructure beyond the means of many middle- and low-income countries. High-throughput, low-technology prevention strategies and point-of-care testing are needed to identify and treat women bearing high-risk cervical dysplasia to prevent cancer development.

A primary prevention strategy relies on HPV vaccines. Currently approved vaccines include the recombinant proteins to the late proteins, L1 and L2 of HPV16 and 18, as well as other, less common cancer-causing isotypes 11, 31, 33, 45, 52, and 58. Vaccination of girls aged 11–13 years with two injections (1 year apart) before the initiation of sexual activity dramatically reduces the rate of high-risk HPV infection and subsequent dysplasia. Vaccination of both boys and girls is increasingly considered to reduce the risk of HPV-induced cancers of the pharynx. Partial protection is also provided against other HPV types, although vaccinated women are still at risk for HPV infection and still benefit from standard Pap smear screening.

CLINICAL PRESENTATIONS

Risk Factors Clinical risk factors include many HPV infection-linked features: a high number of sexual partners, early age of first intercourse, and history of venereal disease. Smoking is a cofactor;

heavy smokers have a higher risk of dysplasia with HPV infection. HIV infection, especially when associated with low CD4+ T-cell counts, is associated with a higher rate of high-grade dysplasia and likely a shorter latency period between infection and invasive disease. Histologically, the majority of cervical malignancies are squamous cell carcinomas associated with HPV, but adenocarcinomas are also HPV related, and both arise in the transitional zone of the endocervical canal; the lesions in the canal or cervical glands may not be seen by visual inspection of the cervix and can be missed by Pap smear screening. Less common malignancies, such as vulvar cancer, anal cancer, and pharyngeal cancer, are also linked to HPV infection.

Diagnosis of Cervical Cancer Early cancer of the cervix is asymptomatic, and this biology underlies the recommendations for routine gynecologic care. Larger, invasive carcinomas often have symptoms or signs including postcoital spotting or intermenstrual cycle bleeding or menometrorrhagia. Foul-smelling or persistent yellow discharge may also be present. Symptoms such as pelvic or sacral pain suggest lateral extension into the pelvic nerve plexus by either the primary tumor or a pelvic node metastasis and indicate advanced-stage disease. Likewise, flank pain from hydronephrosis from ureteral compression or deep-venous thrombosis from iliac vessel compression suggests either extensive nodal disease or direct extension of the primary tumor to the pelvic sidewall. The most common finding upon physical exam is a visible tumor on the cervix, but deeper tumors in the cervical os and glands should be considered. Larger tumors may be identified by inspection and biopsied directly. Staging of cervical cancer is performed by clinical exam. Stage I cervical tumors are confined to the cervix, whereas stage II tumors extend into the upper vagina or paracervical soft tissue (Fig. 89-2). Stage III tumors extend to the lower vagina or the pelvic sidewalls, whereas stage IV tumors invade the bladder or rectum or have spread to distant sites. While radiographic studies are not part of the formal clinical staging of cervical cancer, treatment planning requires them for appropriate therapy. CT can detect hydronephrosis indicative of pelvic sidewall disease but is not accurate at evaluating other pelvic structures. MRI is more accurate at estimating uterine extension and paracervical extension of disease into soft tissues typically bordered by broad and cardinal ligaments that support the uterus in the central pelvis. Very small stage I cervical tumors can be treated with a variety of surgical procedures, but minimally invasive surgery has inferior outcome compared to standard open hysterectomy. In young women desiring to maintain fertility,

radical trachelectomy removes the cervix with subsequent anastomosis of the upper vagina to the uterine corpus; however, subsequent pregnancies may be more problematic. Patients with large stage I cervical tumors (4 cm) confined to the cervix and all stage II to IV patients are treated with radiation therapy in combination with cisplatin-based chemotherapy. This multimodality treatment can offer the patient with advanced-stage disease a 40–80% chance of cure depending on the clinical circumstances. Platinum agents (cisplatin or carboplatin) combined with paclitaxel and bevacizumab are generally considered as the best palliative choice for metastatic cervical cancer patients. Secondary chemotherapy confers minimal improvement in most patients. Immunotherapy with immune checkpoint inhibitors or adoptive T-cell therapies are promising avenues for improved outcomes in recurrent, unresectable cancers of the cervix.

UTERINE CANCER

■ EPIDEMIOLOGY

Several different tumor types arise in the uterine corpus. Most tumors arise in the glandular lining and are endometrial adenocarcinomas. Benign (leiomyomas) and malignant tumors (leiomyosarcomas) can also arise in the uterine smooth muscle and have very different clinical features. The endometrioid histologic subtype is the most common gynecologic malignancy in the United States. In 2021, the American Cancer Society predicted that 66,570 new cancers of the uterine corpus in 2021 with 12,940 resulting deaths. Development of these tumors is a multistep process, with estrogen playing an important early role in driving endometrial gland proliferation. Relative overexposure to this class of hormones is the principal risk factor for the subsequent development of endometrioid tumors. In contrast, progestins drive glandular maturation and are protective. Hence, women with high endogenous or pharmacologic exposure to estrogens, especially if unopposed by progesterone, are at higher risk for endometrial cancer. Obese women, women treated with postmenopausal estrogens, or women with estrogen-producing tumors are at higher risk for endometrial cancer. In addition, long-term treatment with tamoxifen, which has antiestrogenic effects in breast tissue but can show weak estrogenic effects in uterine epithelium, is associated with an increased risk of endometrial cancer.

Genetics Women with a germline mutation in one of a series of DNA mismatch repair genes associated with the Lynch syndrome, also known as hereditary nonpolyposis colon cancer (HNPCC) syndrome, are at increased risk for endometrioid endometrial carcinoma. These individuals have germline mutations in *MSH2*, *MLH1*, and, in rare cases, *PMS1* and *PMS2*.

Individuals who carry these mutations typically have a family history of cancer and are at markedly increased risk for colon cancer and modestly increased risk for ovarian cancer and a variety of other tumors. Middle-aged women with HNPCC carry a 4% annual risk of endometrial cancer and a relative overall risk of ~200-fold as compared to age-matched women without HNPCC. In sporadic cancers, secondary events such as mutation of the *PI3K* gene or the loss of the *PTEN* tumor-suppressor gene likely serve as secondary genetic “hits” in the carcinogenesis related to estrogenic excess. The molecular events that underlie less common endometrial cancers such as clear cell and papillary serous tumors of the uterine corpus are not well understood.

Staging of cervix cancer

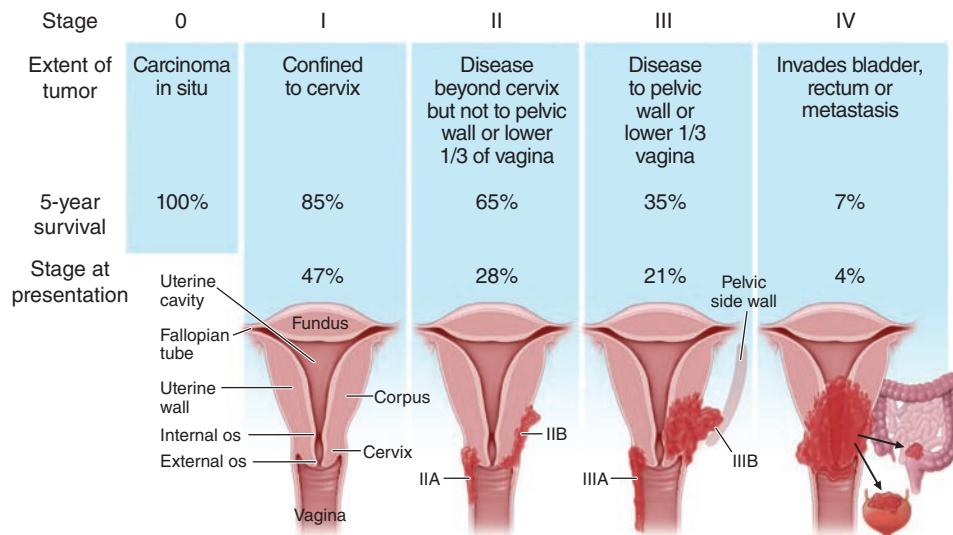


FIGURE 89-2 Anatomic display of the stages of cervix cancer defined by location, extent of tumor, frequency of presentation, and 5-year survival.

■ PATHOLOGY

Approximately 75–80% of endometrial cancers are adenocarcinomas and have been characterized as type 1 (estrogen-linked) endometrial cancers and type 2 cancers that have less clear associations with estrogens (clear cell cancers, serous cancers, and mucinous cancers). Endometrial serous cancers show *TP53* loss of function and behave clinically more like ovarian cancers with high risk for systemic recurrence. Prognosis for endometrial cancer depends on stage, histologic grade, and depth of myometrial invasion.

■ CLINICAL PRESENTATION

The majority of women with tumors of the uterine corpus present with postmenopausal vaginal bleeding due to shedding of the malignant endometrial lining. Premenopausal women often will present with atypical bleeding between typical menstrual cycles. These signs typically bring a woman to the attention of a health care professional, and the majority of women present with early-stage disease in which the tumor is confined to the uterine corpus and, consequently, have a high cure rate. Diagnosis is typically established by endometrial biopsy. Type 1 tumors may spread to pelvic or para-aortic lymph nodes and are generally subjected to sentinel lymph node biopsy at the time of primary surgery. Serous tumors tend to have patterns of spread much more reminiscent of ovarian cancer, and patients may present with omental/peritoneal disease and sometimes ascites. Some women with endometrial cancer have a history of endometriosis. Some women presenting with uterine sarcomas will present with pelvic pain. Uterine sarcomas (carcinosarcomas and leiomyosarcomas) commonly are found by detection of symptomatic large pelvic masses that may or may not be associated with dysfunctional bleeding.

TREATMENT

Uterine Cancer

Most women with endometrial cancer have disease that is localized to the uterus (75% are stage I, Table 89-1), and definitive treatment typically involves a hysterectomy with removal of the ovaries and fallopian tubes. The resection of lymph nodes does not improve outcome, but sentinel node resection provides important staging and prognostic information. Node involvement defines stage IIIC disease. Tumor grade and depth of invasion are the two key prognostic variables in early-stage tumors, and women with low-grade and/or minimally invasive tumors (<50% myometrial penetration) are typically observed after definitive surgical therapy. Patients with high-grade tumors or tumors that are deeply invasive (stage IB) are at higher risk for pelvic recurrence or recurrence at the vaginal cuff, which is typically prevented by intravaginal brachytherapy.

The loss of one or more mismatch repair proteins results in microsatellite instability (MSI) with a larger number of mutations in the tumor. MSI testing should be routinely performed in endometrial cancers at the time of diagnosis to help with current and future treatment plans. MSI cancers, when recurrent or present at an advanced stage, are likely to respond to immune checkpoint therapy. Women with regional metastases or metastatic disease (3% of patients) with low-grade tumors can be treated with progesterone or tamoxifen. Poorly differentiated tumors lack hormone receptors and are typically resistant to hormonal manipulation. The role of adjuvant chemotherapy in stage I-II disease is generally restricted to serous endometrial cancers. For more advanced-stage cancers (stage III-IV), chemotherapy and/or immune checkpoint blockade are administered because of the higher rates of recurrent systemic disease. Carboplatin and paclitaxel combinations are the current standard of care. Chemotherapy for metastatic disease is delivered with palliative intent. Patients with advanced cancer and known mismatch repair deficits may respond particularly well to immunotherapy with antagonists of the PD-1/PD-L1 axis. Lenvatinib and pembrolizumab (even for microsatellite-stable tumors)

have become the most common second-line treatments, although survival data are not yet available. Other potentially active treatments include bevacizumab, mTOR inhibitors (e.g., temsirolimus), and anthracyclines. Carcinosarcomas of the uterus (also called Müllerian tumors) contain both mesenchymal and epithelial components but will often respond to paclitaxel and platinum complex therapy. Other uterine sarcomas require an entirely different approach and need histology-specific consideration. The most common are the leiomyosarcomas of the uterus, which are treated with docetaxel/gemcitabine at recurrence but do not appear to benefit from adjuvant therapy. Ifosfamide/doxorubicin and trabectedin can have some benefit in refractory disease.

■ GESTATIONAL TROPHOBlastic TUMORS

Gestational trophoblastic diseases represent a spectrum of neoplasia from benign hydatidiform mole to choriocarcinoma due to persistent trophoblastic disease associated most commonly with molar pregnancy but occasionally seen after normal gestation. The most common presentations of trophoblastic tumors are partial and complete molar pregnancies. These represent approximately 1 in 1500 conceptions in developed Western countries. The incidence widely varies globally, with areas in Southeast Asia having a much higher incidence of molar pregnancy. Regions with high molar pregnancy rates are often associated with diets low in carotene and animal fats.

■ RISK FACTORS

Trophoblastic tumors result from the outgrowth or persistence of placental tissue. They arise most commonly in the uterus but can also arise in other sites such as the fallopian tubes due to ectopic pregnancy. Risk factors include poorly defined dietary and environmental factors as well as conceptions at the extremes of reproductive age, with the incidence particularly high in females conceiving at younger than age 16 or older than age 50. In older women, the incidence of molar pregnancy might be as high as one in three, likely due to increased risk of abnormal fertilization of the aged ova. Most trophoblastic neoplasms are associated with complete moles, diploid tumors with all genetic material from the paternal donor (known as uniparental disomy). This is thought to occur when a single sperm fertilizes an enucleate egg that subsequently duplicates the paternal DNA. Trophoblastic proliferation occurs with exuberant villous stroma. If pseudopregnancy extends out past the 12th week, fluid progressively accumulates within the stroma, leading to “hydropic changes.” Fetal development does not occur in complete moles.

Partial moles arise from the fertilization of an egg with two sperm cells; hence, two-thirds of genetic material is paternal in these triploid tumors. Hydropic changes are less dramatic, and fetal development can often occur through late first trimester or early second trimester, at which point spontaneous abortion is common. Laboratory findings will include excessively high human chorionic gonadotropin (hCG) and high AFP. The risk of persistent gestational trophoblastic disease after partial mole is ~5%. Complete and partial moles can be noninvasive or invasive. Myometrial invasion occurs in no more than one in six complete moles and a lower portion of partial moles.

■ PRESENTATION OF INVASIVE TROPHOBlastic DISEASE

The clinical presentation of molar pregnancy is changing in developed countries due to the early detection of pregnancy with home pregnancy kits and the very early use of Doppler and ultrasound to evaluate the early fetus and uterine cavity for evidence of a viable fetus. Thus, in these countries, the majority of women presenting with trophoblastic disease have their moles detected early and have typical symptoms of early pregnancy including nausea, amenorrhea, and breast tenderness. With uterine evacuation of early complete and partial moles, most women experience spontaneous remission of their disease as monitored by serial serum β-hCG levels. These women require no

chemotherapy. Patients with persistent elevation of β -hCG or rising β -hCG after uterine evacuation have persistent or actively growing gestational trophoblastic disease and require therapy. Most series suggest that between 15 and 25% of women will have evidence of persistent gestational trophoblastic disease after molar evacuation.

In women who lack access to prenatal care, presenting symptoms can be life-threatening, including the development of preeclampsia or even eclampsia. Hyperthyroidism can also be seen with very high β -hCG values. Evacuation of large moles can be associated with life-threatening complications including uterine perforation, volume loss, high-output cardiac failure, and adult respiratory distress syndrome (ARDS).

For women with evidence of rising β -hCG or radiologic confirmation of metastatic or persistent regional disease, prognosis can be estimated through a variety of scoring algorithms that identify women at low, intermediate, and high risk for requiring multiagent chemotherapy. In general, women with widely metastatic nonpulmonary disease, very elevated β -hCG, and prior normal antecedent term pregnancy are considered at high risk and typically require multiagent chemotherapy at an expert center for cure. Even very advanced gestational trophoblastic disease is almost uniformly curable when managed by an expert in this rare malignancy.

TREATMENT

Invasive Trophoblastic Disease

Management of invasive trophoblastic disease should be 100% curative, and complex patients should be managed by clinicians experienced in this disease. The management for a persistent and rising β -hCG after evacuation of a molar conception is typically chemotherapy, although surgery can play an important role for chemotherapy-resistant disease that is isolated in the uterus (especially if childbearing is complete) or to control hemorrhage. For women wishing to maintain fertility or with metastatic disease, the preferred treatment is chemotherapy. Trophoblastic disease is exquisitely sensitive to chemotherapy, and guided by serial serum β -hCG testing, successful, curative treatment is the rule. Single-agent treatment with dactinomycin or methotrexate cures 90% of women with low-risk disease. Patients with high-risk disease (very high β -hCG levels, presentation ≥ 4 months after pregnancy, brain or liver metastases, failure of methotrexate therapy) are typically treated with multiagent chemotherapy (etoposide, methotrexate, and dactinomycin, alternating with cyclophosphamide and vincristine [EMA-CO]), which is typically curative even in women with extensive metastatic disease. A regimen of cisplatin and etoposide alternating with etoposide/methotrexate/dactinomycin is used for the highest-risk patients. In the highest-risk patients with liver, lung, and brain metastases, hemorrhage from the rich tumor vasculature is a major risk during chemotherapy initiation. Cured women may become pregnant again without evidence of increased fetal or maternal complications.

FURTHER READING

- GREEN AK et al: A review of immune checkpoint blockade therapy in endometrial cancer. *Am Soc Clin Oncol Educ Book* 40:1, 2020.
- LONGO DL: Personalized medicine for primary treatment of serous ovarian cancer. *N Engl J Med* 381:2471, 2019.
- LU KH, BROADDUS RR: Endometrial cancer. *N Engl J Med* 383:2053, 2020.
- MENON U et al: Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet* 397:2182, 2021.
- RAMIREZ PT et al: Minimally invasive versus abdominal radical hysterectomy for cervical cancer. *N Engl J Med* 379:1895, 2018.

90

Primary and Metastatic Tumors of the Nervous System

Lisa M. DeAngelis, Patrick Y. Wen

An estimated 87,000 people will be diagnosed with a primary brain tumor annually in the United States. At least 25,000 of these tumors are malignant, and most of these are gliomas. Meningiomas account for 35%, vestibular schwannomas 10%, and central nervous system (CNS) lymphomas ~2%. Brain metastases are three times more common than all primary brain tumors combined and are diagnosed in ~150,000 people each year. Metastases to the leptomeninges and epidural space of the spinal cord each occur in ~3–5% of patients with systemic cancer and are also a major cause of neurologic disability.

APPROACH TO THE PATIENT

Primary and Metastatic Tumors of the Nervous System

CLINICAL FEATURES

Brain tumors of any type can present with a variety of symptoms and signs that fall into two categories: general and focal; patients often have a combination of the two (Table 90-1). General symptoms include headache, with or without nausea or vomiting, cognitive difficulties, personality change, and gait disorder. These symptoms arise when the enlarging tumor and its surrounding edema cause an increase in intracranial pressure or compression of cerebrospinal fluid (CSF) circulation leading to hydrocephalus. The classic brain tumor headache predominates in the morning and improves during the day, but this pattern is seen in a minority of patients. Headaches are often holocephalic but can be ipsilateral to the side of a tumor. Occasionally, headaches have features of a typical migraine with unilateral throbbing pain associated with visual scotoma. Personality changes may include apathy and withdrawal from social situations, mimicking depression. Focal or lateralizing findings include hemiparesis, aphasia, or visual field defect. Lateralizing symptoms are typically subacute and progressive; language difficulties may be mistaken for confusion. Seizures are common, occurring in ~25% of patients with brain metastases or malignant gliomas, and are the presenting symptom in up to 90% of patients with a low-grade glioma. All seizures arising from a brain tumor will have a focal onset whether or not it is apparent clinically.

NEUROIMAGING

Cranial magnetic resonance imaging (MRI) is the preferred diagnostic test for any patient suspected of having a brain tumor and should be performed with gadolinium contrast administration. Computed tomography (CT) scan should be reserved for those patients unable to undergo MRI. Malignant brain tumors—whether primary or metastatic—typically enhance with gadolinium, have central areas of necrosis, and are surrounded by edema of the neighboring white matter. Low-grade gliomas usually do not enhance with gadolinium and are best appreciated on fluid-attenuated inversion recovery (FLAIR) MRI sequences. Meningiomas have a typical appearance on MRI because they are dural-based enhancing tumors with a dural tail and compress but do not invade the brain. Dural metastases or a dural lymphoma can have a similar appearance. Imaging is characteristic for many primary and metastatic tumors and sometimes will suffice to establish a diagnosis when the location precludes surgical intervention (e.g., brainstem glioma).

TABLE 90-1 Symptoms and Signs at Presentation of Brain Tumors

	HIGH-GRADE GLIOMA (%)	LOW-GRADE GLIOMA (%)	MENINGIOMA (%)	METASTASES (%)
Generalized				
Impaired cognitive function	50	10	30	60
Hemiparesis	40	10	36	60
Headache	50	40	37	50
Lateralizing				
Seizures	20	70+	17	18
Aphasia	20	<5	—	18
Visual field deficit	—	—	—	7

Functional MRI is useful in presurgical planning to define eloquent sensory, motor, or language cortex. Positron emission tomography (PET) is useful in determining the metabolic activity of the lesions seen on MRI; MR perfusion and spectroscopy can provide information on blood flow or tissue composition. These techniques may help distinguish tumor progression from tissue necrosis due to treatment with radiation and chemotherapy. Neuroimaging is the only test necessary to diagnose a brain tumor. Laboratory tests are rarely useful, although patients with metastatic disease may have elevation of a serum tumor marker (e.g., β human chorionic gonadotropin [β -hCG] from testicular cancer). Additional testing such as cerebral angiogram, electroencephalogram (EEG), or lumbar puncture is rarely indicated or helpful.

TREATMENT

Brain Tumors

Therapy of any intracranial malignancy requires both symptomatic and definitive treatments. Definitive treatment is based on the specific tumor type and includes surgery, radiotherapy, and chemotherapy. However, symptomatic treatments apply to brain tumors of any type. Most high-grade malignancies are accompanied by substantial surrounding edema, which contributes to neurologic disability and raised intracranial pressure. Glucocorticoids are highly effective at reducing perilesional edema and improving neurologic function, often within hours of administration. Dexamethasone has been the glucocorticoid of choice because of its relatively low mineralocorticoid activity; initial doses are 8–12 mg/d. Glucocorticoids rapidly ameliorate symptoms and signs, but their long-term use causes substantial toxicity including insomnia, weight gain, diabetes mellitus, steroid myopathy, and personality changes. Consequently, a taper is indicated as definitive treatment is administered and the patient improves.

Patients with brain tumors who present with seizures require antiepileptic drug therapy. Prophylactic antiepileptic drugs are used in the perioperative setting, but there is no role for extended use in patients who have not had a seizure. The agents of choice are those drugs that do not induce the hepatic microsomal enzyme system. These include levetiracetam, topiramate, lamotrigine, valproic acid, and lacosamide ([Chap. 425](#)). Other drugs, such as phenytoin and carbamazepine, are used less frequently because they are potent enzyme inducers that can interfere with both glucocorticoid and chemotherapy metabolism. Venous thromboembolic disease occurs in 20–30% of patients with high-grade gliomas or brain metastases. Prophylactic anticoagulants should be used during hospitalization and in nonambulatory patients. Those who have had either a deep vein thrombosis or a pulmonary embolus can receive therapeutic doses of anticoagulation safely and without increasing the risk for hemorrhage into the tumor. Inferior vena cava filters are reserved for patients with absolute contraindications to anticoagulation such as recent craniotomy.

PRIMARY BRAIN TUMORS

■ EPIDEMIOLOGY

No etiology has been identified for most primary brain tumors. The only established risk factors are exposure to ionizing radiation (meningiomas, gliomas, and schwannomas) and immunosuppression (primary CNS lymphoma). There is no proven evidence for any association with exposure to electromagnetic fields including cellular telephones, head injury, foods containing *N*-nitroso compounds, or occupational risk factors. A small minority of patients have a family history of brain tumors. Some of these familial cases are associated with genetic syndromes ([Table 90-2](#)).

■ MOLECULAR PATHOGENESIS

As with other neoplasms, brain tumors arise as a result of a multistep process driven by the sequential acquisition of genetic alterations. These include loss of tumor-suppressor genes (e.g., *p53*, cyclin-dependent kinase inhibitor 2A and 2B [*CDKN2A/B*], and phosphatase and tensin homolog on chromosome 10 [*PTEN*]) and amplification and overexpression of protooncogenes such as the epidermal growth factor receptor (*EGFR*) and platelet-derived growth factor receptors (*PDGFR*). The accumulation of these genetic abnormalities results in uncontrolled cell growth and tumor formation. Many brain tumors, including glioblastomas, are characterized by significant molecular heterogeneity, which contributes to the difficulty in developing effective therapies.

Important progress has been made in understanding the molecular pathogenesis of several types of brain tumors, including glioblastoma and medulloblastoma, allowing them to be separated into different subtypes with different prognoses. This has led the World Health Organization (WHO) to issue an update on the classification of CNS tumors in 2016 that for the first time incorporates molecular parameters in addition to traditional histology into the diagnosis of brain tumors.

INTRINSIC “MALIGNANT” TUMORS

■ DIFFUSE GLIOMAS

Gliomas are the most common type of malignant primary brain tumor and are derived, based on their presumed lineage, into astrocytomas and oligodendrogliomas. These tumors are classified based on two highly recurrent molecular alterations, isocitrate dehydrogenase (*IDH*) mutations and 1p/19q codeletion, in addition to more conventional histopathologic parameters. Most lower-grade astrocytomas have *IDH* mutations but intact 1p/19q and often mutations in *ATRX* and *p53*. Oligodendrogliomas usually have *IDH* mutations and codeletion of 1p/19q. A minority of astrocytomas and oligodendrogliomas that lack *IDH* mutations (20–30%) have a worse prognosis.

Diffuse gliomas can present rarely as widespread infiltration of the brain tissue without a focal mass. Such tumors usually present with cognitive problems, and the MRI demonstrates confluent, typically nonenhancing areas of increased signal on FLAIR sequences without significant mass effect. Formerly known as gliomatosis cerebri, these lesions are now categorized by the pathology identified on biopsy, but they can be diagnostically challenging when the nature of the imaging

TABLE 90-2 Genetic Syndromes Associated with Primary Brain Tumors

SYNDROME	INHERITANCE	GENE/PROTEIN	ASSOCIATED TUMORS
Cowden's syndrome	AD	Mutations of <i>PTEN</i> (ch10p23)	Dysplastic cerebellar gangliocytoma (Lhermitte-Duclos disease), meningioma, astrocytoma Breast, endometrial, thyroid cancer, trichilemmomas
Familial schwannomatosis	Sporadic Hereditary	Mutations in <i>INI1/SNF5</i> (ch22q11)	Schwannomas, gliomas
Gardner's syndrome	AD	Mutations in <i>APC</i> (ch5q21)	Medulloblastoma, glioblastoma, craniopharyngioma Familial polyposis, multiple osteomas, skin and soft tissue tumors
Gorlin syndrome (basal cell nevus syndrome)	AD	Mutations in <i>Patched 1</i> gene (ch9q22.3)	Medulloblastomas Basal cell carcinoma
Li-Fraumeni syndrome	AD	Mutations in <i>p53</i> (ch17p13.1)	Gliomas, medulloblastomas Sarcomas, breast cancer, leukemias, others
Lynch syndrome	AD	Mutations in <i>MSH2</i> , <i>MSH1</i> , <i>MSH6</i> , <i>PMS2</i>	Glioblastoma and other gliomas Gastrointestinal, endometrial, and other cancers
Multiple endocrine neoplasia 1 (Wermer's syndrome)	AD	Mutations in <i>Menin</i> (ch11q13)	Pituitary adenoma, malignant schwannomas Parathyroid and pancreatic islet cell tumors
NF1	AD	Mutations in <i>NF1</i> /neurofibromin (ch17q12-22)	Schwannomas, astrocytomas, optic nerve gliomas, meningiomas Neurofibromas, neurofibrosarcomas, others
NF2	AD	Mutations in <i>NF2</i> /merlin (ch22q12)	Bilateral vestibular schwannomas, astrocytomas, multiple meningiomas, ependymomas
TSC (Bourneville disease)	AD	Mutations in <i>TSC1/TSC2</i> (ch9q34/16)	Subependymal giant cell astrocytoma, ependymomas, glioma, ganglioneuroma, hamartoma
Turcot syndrome	AD AR	Mutations in <i>APC</i> ^a (ch5) <i>hMLH1</i> (ch3p21)	Gliomas, medulloblastomas Adenomatous colon polyps, adenocarcinoma
VHL	AD	Mutations in <i>VHL</i> gene (ch3p25)	Hemangioblastomas Retinal angiomas, renal cell carcinoma, pheochromocytoma, pancreatic tumors and cysts, endolymphatic sac tumors of the middle ear

^aVarious DNA mismatch repair gene mutations may cause a similar clinical phenotype, also referred to as Turcot syndrome, in which there is a predisposition to nonpolyposis colon cancer and brain tumors.

Abbreviations: AD, autosomal dominant; APC, adenomatous polyposis coli; AR, autosomal recessive; ch, chromosome; NF, neurofibromatosis; PTEN, phosphatase and tensin homologue; TSC, tuberous sclerosis complex; VHL, von Hippel-Lindau.

abnormalities is unclear. Often diagnosis is delayed until the patient develops worsening deficits or there is clear progression on imaging. Treatment is then determined by the pathology.

■ ASTROCYTOMAS

These are infiltrative tumors with a presumptive glial cell of origin. The WHO classifies astrocytomas into four prognostic grades based on histologic features: grade I (pilocytic astrocytoma, subependymal giant cell astrocytoma); grade II (astrocytoma); grade III (anaplastic astrocytoma); and grade IV (glioblastoma). Grades I and II are considered low-grade, and grades III and IV high-grade, astrocytomas.

Low-Grade Astrocytoma • GRADE I ASTROCYTOMAS Pilocytic astrocytomas (WHO grade I) are the most common tumor of childhood. They occur typically in the cerebellum but may also be found elsewhere in the neuraxis, including the optic nerves and brainstem. Frequently they appear as cystic lesions with an enhancing mural nodule. Often they have *BRAF* fusions or mutations. These are well-demarcated lesions that are potentially curable if they can be resected completely. Giant cell subependymal astrocytomas are usually found in the ventricular wall of patients with tuberous sclerosis. They often do not require intervention but can be treated surgically or with inhibitors of the mammalian target of rapamycin (mTOR).

GRADE II ASTROCYTOMAS These are infiltrative tumors that usually present with seizures in young adults. They appear as nonenhancing tumors with increased T2/FLAIR signal (Fig. 90-1). If feasible, patients should undergo maximal surgical resection, although complete resection is rarely possible because of the invasive nature of the tumor. In patients at higher risk for recurrence (subtotal resection or above the age of 40 years), there is evidence that radiation therapy (RT) followed by PCV (procarbazine, cyclohexylchloroethylnitrosourea [CCNU], and vincristine) chemotherapy may possibly be of benefit. The tumor

transforms to a malignant astrocytoma in most patients, leading to variable survival with a median of ~5–10 years.

High-Grade Astrocytoma • GRADE III (ANAPLASTIC) ASTROCYTOMA These account for ~15–20% of high-grade

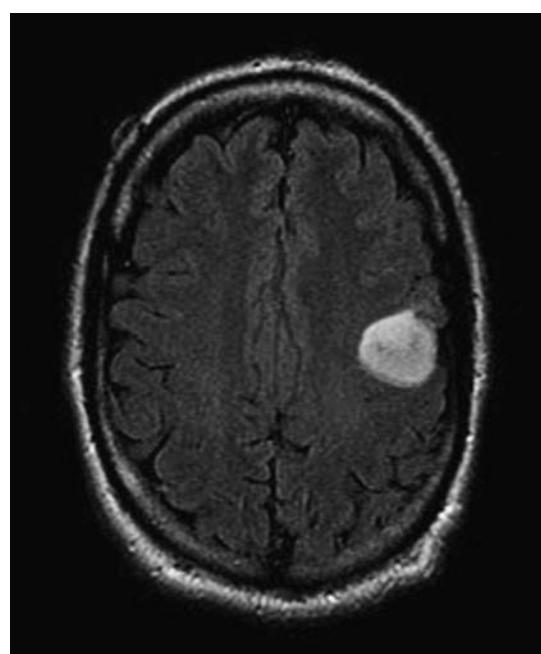


FIGURE 90-1 Fluid-attenuated inversion recovery (FLAIR) MRI of a left frontal low-grade astrocytoma. This lesion did not enhance.

astrocytomas. They generally present in the fourth and fifth decades of life as variably enhancing tumors. Treatment is the same as for glioblastoma, consisting of maximal safe surgical resection followed by RT and adjuvant temozolomide alone or RT with concurrent and adjuvant temozolomide.

GRADE IV ASTROCYTOMA (GLIOBLASTOMA) Glioblastoma accounts for the majority of high-grade astrocytomas. Approximately 10% of glioblastomas have *IDH* mutations. These tend to arise from lower-grade tumors (secondary glioblastomas) and have a better prognosis. In the next update of the WHO classification, the term glioblastoma will be restricted to tumors without *IDH* mutations. Glioblastomas with *IDH* mutations have a different biology and better prognosis and will be termed astrocytomas, *IDH*-mutant, grade 4. In addition, astrocytomas without the classic histologic features of glioblastoma (necrosis and endothelial proliferation) but that have the molecular features of glioblastoma (epidermal growth factor amplification, combined with whole chromosome 7 gain and 10 loss, or telomerase reverse transcriptase [*TERT*] promoter mutations) will also be considered glioblastomas.

Glioblastomas are the most common malignant primary brain tumor, with >12,000 cases diagnosed each year in the United States. Patients usually present in the sixth and seventh decades of life with headache, seizures, or focal neurologic deficits. The tumors appear as ring-enhancing masses with central necrosis and surrounding edema (Fig. 90-2). These are highly infiltrative tumors, and the areas of increased T2/FLAIR signal surrounding the main tumor mass contain invading tumor cells. Treatment involves maximal surgical resection followed by partial-field external-beam RT (6000 cGy in thirty 200-cGy fractions) with concomitant temozolomide, followed by 6 months of adjuvant temozolomide. With this regimen, median survival is increased to 14.6–18 months compared to only 12 months with RT alone, and 5-year survival is ~10%. Efforts to increase the dose of RT locally using brachytherapy or stereotactic radiosurgery (SRS) have failed to improve the outcome and these treatments are not used. Patients whose tumor contains the DNA repair enzyme O⁶-methylguanine-DNA methyltransferase (MGMT) are relatively resistant to temozolomide and have a worse prognosis compared to those whose tumors contain low levels of MGMT as a result of silencing of the *MGMT* gene by promoter hypermethylation. Implantation of biodegradable polymers containing carmustine chemotherapy into the tumor bed after resection of the tumor or addition of tumor treating fields (scalp electrodes delivering low-intensity electric currents) produces a modest improvement in survival.

For elderly patients aged >65–70 years, a hypofractionated RT regimen of 40 Gy over 3 weeks with temozolomide is well tolerated and likely leads to similar outcomes as the 6-week standard RT regimen.

Despite optimal therapy, glioblastomas invariably recur. Treatment options for recurrent disease may include reoperation, re-irradiation with bevacizumab, and alternate chemotherapeutic regimens. Bevacizumab, a humanized vascular endothelial growth factor (VEGF) monoclonal antibody, has activity in recurrent glioblastoma, increasing progression-free survival but not overall survival and reducing peritumoral edema and glucocorticoid use (Fig. 90-3). Immune checkpoint inhibitors have been successful in a variety of solid tumors but have failed to demonstrate substantial activity in glioblastoma. A recent phase III trial comparing bevacizumab with nivolumab in recurrent glioblastoma demonstrated an identical median overall survival of 9.8–10 months in the two arms, with similar toxicities. Treatment decisions for patients with recurrent glioblastoma must be made on an individual basis, taking into consideration such factors as previous therapy, time to relapse, performance status, and quality of life. Whenever feasible, patients should be enrolled in clinical trials. Novel therapies

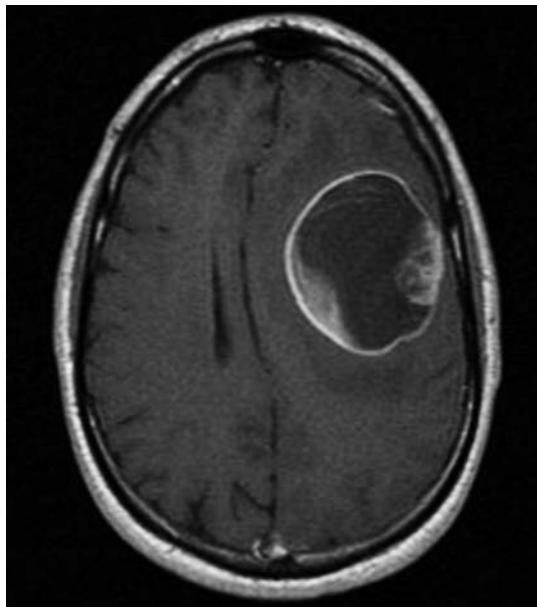
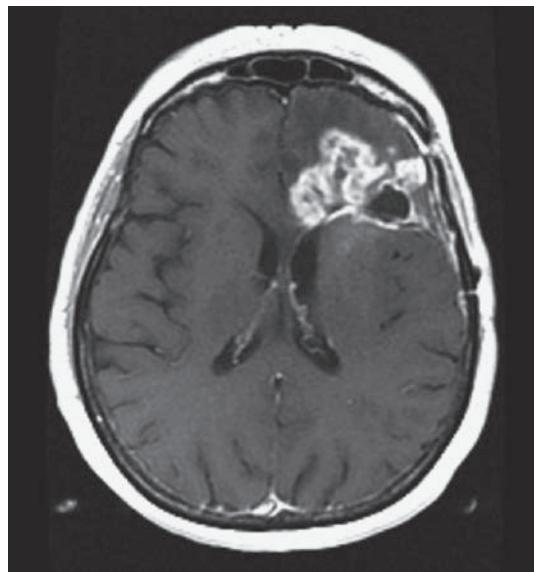


FIGURE 90-2 Postgadolinium T1 MRI of a large cystic left frontal glioblastoma.



A

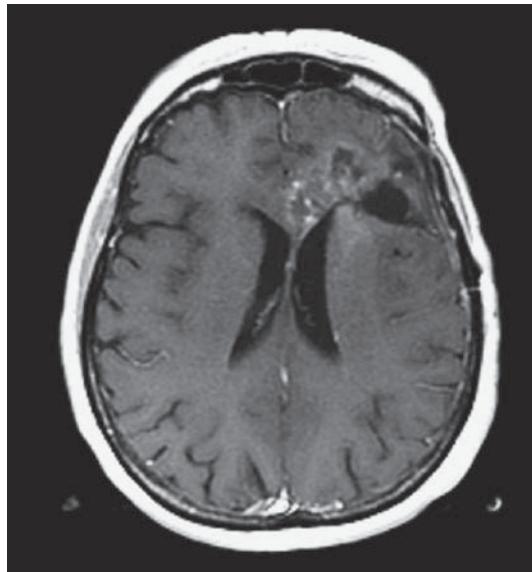


FIGURE 90-3 Postgadolinium T1 MRI of a recurrent glioblastoma before (A) and after (B) administration of bevacizumab. Note the decreased enhancement and mass effect.

undergoing evaluation in patients with glioblastoma include targeted molecular agents directed at receptor tyrosine kinases and signal transduction pathways; immunotherapy using vaccines, novel checkpoint inhibitors, or chimeric antigen receptor (CAR) T cells; oncolytic viruses; antiangiogenic agents; chemotherapeutic agents that cross the blood-brain barrier more effectively than currently available drugs; and infusion of radiolabeled drugs and targeted toxins into the tumor and surrounding brain by means of convection-enhanced delivery.

The most important adverse prognostic factors in patients with glioblastomas are older age, absence of *IDH* mutations, unmethylated MGMT promoter, poor Karnofsky performance status, and unresectable tumor.

Gliosarcomas are a variant of glioblastoma containing both an astrocytic and a sarcomatous component and are treated in the same way as glioblastomas.

■ OLIGODENDROGLIOMA

Oligodendrogiomas account for ~15–20% of gliomas. They are characterized by codeletion of 1p/19q and have *IDH* mutations. Oligodendrogiomas are classified by the WHO into oligodendrogiomas (grade II) or anaplastic oligodendrogiomas (AOs) (grade III). Oligodendrogiomas have distinctive pathologic features such as perinuclear clearing—giving rise to a “fried-egg” appearance—and a reticular pattern of blood vessel growth. Some tumors have both an oligodendroglial as well as an astrocytic component. With molecular testing, it is now clear that almost all of these mixed tumors (oligoastrocytomas) are genetically either astrocytomas or oligodendrogiomas. As a result, the diagnosis of oligoastrocytoma is now rarely made unless molecular testing is not available.

Grade II oligodendrogiomas are generally more responsive to therapy and have a better prognosis than pure astrocytic tumors. These tumors present similarly to grade II astrocytomas in young adults. The tumors are nonenhancing and often partially calcified. They should be treated with surgery and, in patients with residual disease or aged >40 years, RT and chemotherapy. Patients with oligodendrogiomas have a median survival in excess of 10 years.

AOs present in the fourth and fifth decades as variably enhancing tumors. They are more responsive to therapy than grade III astrocytomas. Treatment involves maximal safe resection followed by RT and PCV or temozolamide chemotherapy. Median survival of patients with AO is in excess of 10 years.

■ EPENDYOMAS

Ependymomas are tumors derived from ependymal cells that line the ventricular surface. They account for ~5% of childhood tumors, frequently arise from the wall of the fourth ventricle in the posterior fossa, are associated with *RELA* fusions, and are classified as type A and B ependymoma subtypes. Although adults can have intracranial ependymomas, they occur more commonly in the spine, especially in the filum terminale of the spinal cord where they have a myxopapillary histology. Ependymomas that can be completely resected are potentially curable. Partially resected ependymomas will recur and require irradiation. The less common anaplastic ependymoma is more aggressive and is treated with resection and RT; chemotherapy has limited efficacy. Subependymomas are slow-growing benign lesions arising in the wall of ventricles that often do not require treatment.

■ OTHER LESS COMMON GLIOMAS

Gangliogliomas and pleomorphic xanthoastrocytomas occur in young adults. They behave as more indolent forms of grade I gliomas and are usually treated with surgery. Frequently they will have *BRAF* V600E mutations. Brainstem gliomas usually occur in children or young adults and often have *H3K27M* mutations. Despite treatment with RT and chemotherapy, the prognosis is poor, with a median survival of only 1 year.

■ PRIMARY CENTRAL NERVOUS SYSTEM LYMPHOMA

Primary central nervous system lymphoma (PCNSL) is a rare non-Hodgkin’s lymphoma accounting for <3% of primary brain tumors. For

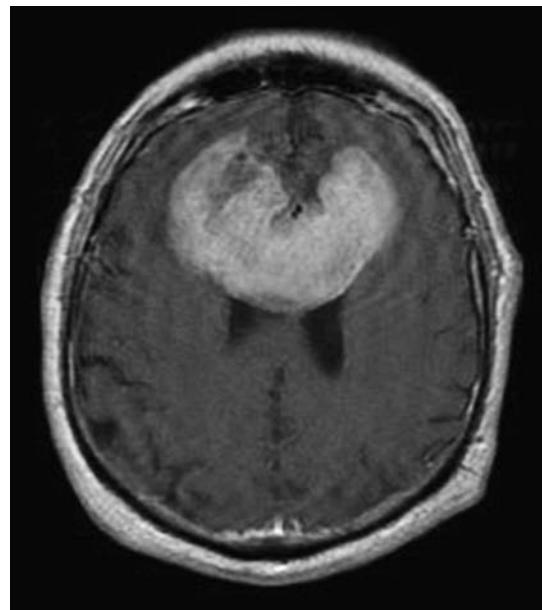


FIGURE 90-4 Postgadolinium T1 MRI demonstrating a large bifrontal primary central nervous system lymphoma (PCNSL). The periventricular location and diffuse enhancement pattern are characteristic of lymphoma.

unclear reasons, its incidence is increasing, particularly in immunocompetent, older individuals.

PCNSL in immunocompetent patients is usually a diffuse large B-cell lymphoma. Immunocompromised patients, especially those infected with the human immunodeficiency virus (HIV) or organ transplant recipients, are at risk for PCNSL that is typically large cell with immunoblastic and more aggressive features. Epstein-Barr virus (EBV) plays an important role in the pathogenesis of PCNSL in this population. These patients are usually severely immunocompromised, with CD4 counts of <50/mL.

Immunocompetent patients with PCNSL are older (median 60 years) than those with HIV-related PCNSL (median 31 years). PCNSL usually presents as a mass lesion, with neuropsychiatric symptoms, lateralizing signs, or seizures. Ocular and leptomeningeal involvement each occur in 15–20% of patients, and involvement of these compartments may be asymptomatic. Rarely, it may present as isolated ocular lymphoma or as primary leptomeningeal lymphoma. When restricted to the leptomeninges, it may present as a subacute or chronic meningitis that causes progressive cranial and spinal nerve dysfunction. CSF cytologic examination or flow cytometry is required to establish the diagnosis.

On contrast-enhanced MRI, PCNSL usually appears as a densely enhancing tumor (Fig. 90-4). Immunocompetent patients have solitary lesions more often than immunosuppressed patients. Frequently there is involvement of the basal ganglia, corpus callosum, or periventricular region. Stereotactic biopsy is necessary to obtain a histologic diagnosis. Whenever possible, glucocorticoids should be withheld until after the biopsy has been obtained because they have a cytolytic effect on lymphoma cells and may lead to nondiagnostic tissue. In addition, patients should be tested for HIV, and the extent of disease should be assessed by performing PET or CT of the body, MRI of the spine, CSF analysis, and slit-lamp examination of the eye. Bone marrow biopsy and testicular ultrasound are occasionally performed.

TREATMENT

Primary Central Nervous System Lymphoma

PCNSL is more sensitive to glucocorticoids, chemotherapy, and RT than other primary brain tumors. Durable complete responses and long-term survival are possible with these treatments. High-dose methotrexate, a folate antagonist that interrupts DNA synthesis, produces response rates ranging from 35 to 80% and median survival

of up to 50 months. The combination of methotrexate with other chemotherapeutic agents such as cytarabine increases the response rate to 70–100%. The addition of whole-brain RT (WBRT) to methotrexate-based chemotherapy prolongs progression-free survival but not overall survival, but it is associated with delayed neurotoxicity, especially in patients aged >60 years. As a result, full-dose RT is frequently omitted, but there may be a role for reduced-dose RT. The anti-CD20 monoclonal antibody rituximab is often incorporated into the chemotherapy regimen, although there are studies questioning its benefit. For some patients, high-dose chemotherapy with autologous stem cell rescue may offer the best chance of preventing relapse. At least 50% of patients will eventually develop recurrent disease. Treatment options include RT for patients who have not had prior irradiation, retreatment with methotrexate, as well as other chemotherapeutic agents such as temozolamide and pemetrexed. High-dose chemotherapy with autologous stem cell rescue may be appropriate in selected patients with relapsed disease. Bruton's tyrosine kinase (BTK) inhibitors such as ibrutinib, immunomodulatory drugs such as pomalidomide and lenalidomide, and immune checkpoint inhibitors have shown promising preliminary activity and are being evaluated in clinical trials, as are CAR T cells.

PCNSL IN IMMUNOCOMPROMISED PATIENTS

PCNSL in immunocompromised patients often produces multiple ring-enhancing lesions that can be difficult to differentiate from metastases or infections such as toxoplasmosis. The diagnosis is usually established by examination of the CSF for cytology and EBV DNA; toxoplasmosis serologic testing; brain PET imaging for hypermetabolism of the lesions, which, although nonspecific, can be consistent with tumor; and, if necessary, brain biopsy. Since the advent of highly active antiretroviral drugs, the incidence of HIV-related PCNSL has declined. These patients are preferably treated with high-dose methotrexate-based regimens and initiation of highly active antiretroviral therapy; WBRT is reserved for those who cannot tolerate systemic chemotherapy. In organ transplant recipients, reduction of immunosuppression may improve outcome.

MEDULLOBLASTOMAS

Medulloblastomas are the most common malignant brain tumor of childhood, accounting for ~20% of all primary CNS tumors among children. They arise from granule cell progenitors or from multipotent progenitors from the ventricular zone. Approximately 5% of children with medulloblastoma have an inherited syndrome, such as Gorlin, Turcot, or Li-Fraumeni, which predisposes to the development of medulloblastoma. Histologically, medulloblastomas are highly cellular tumors with abundant dark staining, round nuclei, and rosette formation (Homer-Wright rosettes). In the 2016 WHO pathologic classification, they have been divided into four molecular subgroups: (1) WNT-activated (primarily affects children and has the best outcome); (2) SHH-activated (affects adults, infants, and children, with the younger patients having the better outcome and adults doing poorly); (3) non-WNT/non-SHH, group 3 (frequently has disseminated CNS disease at diagnosis and has the worst outcome); and (4) non-WNT/non-SHH, group 4 (30% have metastases at diagnosis, but 5-year progression-free survival is 95%). Regardless of subtype, patients present with headache, ataxia, and signs of brainstem involvement. On MRI, they appear as densely enhancing tumors in the posterior fossa, sometimes associated with hydrocephalus. Treatment involves maximal surgical resection, craniospinal irradiation, and chemotherapy with agents such as cisplatin, lomustine, cyclophosphamide, and vincristine. Approximately 70% of patients overall have long-term survival but usually at the cost of significant neurocognitive impairment. A major goal of current research is to improve survival while minimizing long-term complications, and clinical trials are now being designed for specific molecular subgroups.

PINEAL REGION TUMORS

A large variety of tumors can arise in the region of the pineal gland. These typically present with headache, visual symptoms, and hydrocephalus.

Patients may have Parinaud's syndrome characterized by impaired upgaze and accommodation. Some pineal tumors such as pineocytomas and benign teratomas can be treated by surgical resection. Germinomas respond to irradiation, whereas pineoblastomas and nongerminomatous germ cell tumors require craniospinal radiation and chemotherapy.

EXTRINSIC “BENIGN” TUMORS

■ MENINGIOMAS

Meningiomas are diagnosed with increasing frequency as more people undergo neuroimaging for various indications. They are now the most common primary brain tumor, accounting for ~35% of the total. Their incidence increases with age. They tend to be more common in women and in patients with neurofibromatosis type 2 (NF2). They also occur more commonly in patients with a history of cranial irradiation.

Meningiomas arise from the dura mater and are composed of neoplastic meningothelial (arachnoidal cap) cells. They are most commonly located over the cerebral convexities, especially adjacent to the sagittal sinus, but they can also occur in the skull base and along the dorsum of the spinal cord. Meningiomas are classified by the WHO into three histologic grades of increasing aggressiveness: grade I (benign), grade II (atypical), and grade III (malignant).

Many meningiomas are found incidentally following neuroimaging for unrelated reasons. They can also present with headaches, seizures, or focal neurologic deficits. On imaging studies, they have a characteristic appearance usually of a densely enhancing extra-axial tumor arising from the dura (Fig. 90-5). Typically they have a dural tail, consisting of thickened, enhanced dura extending like a tail from the mass. The main differential diagnosis of meningioma is a dural metastasis.

If the meningioma is small and asymptomatic, no intervention is necessary and the lesion can be observed with serial MRI studies. Larger, symptomatic lesions should be resected. If complete resection is achieved, the patient is cured. Incompletely resected tumors tend to recur, although the rate of recurrence can be very slow with grade I tumors. Tumors that cannot be resected, or can only be partially removed, may benefit from external-beam RT or SRS. These treatments may also be helpful in patients whose tumor has recurred after surgery. Hormonal therapy and chemotherapy are currently unproven.

Rarer tumors that resemble meningiomas include hemangiopericytomas and solitary fibrous tumors. Since they share similar molecular alterations (NAB2-STAT6 fusion), the 2016 WHO classification introduced the combined term *solitary fibrous tumor/hemangiopericytoma* for this entity. These tumors are treated with surgery and RT but have a higher propensity to recur locally or metastasize systemically.

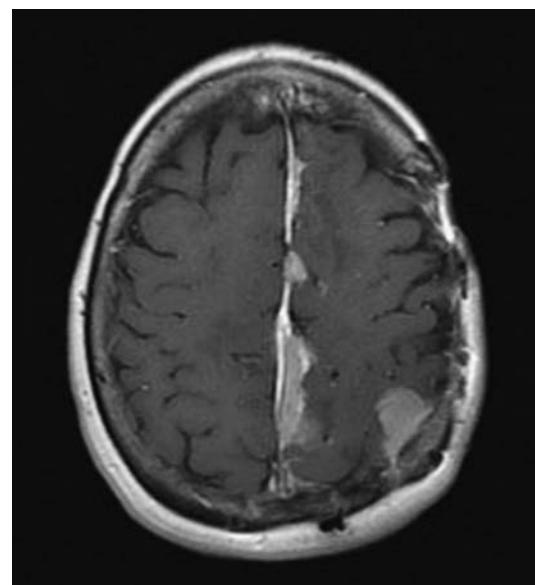


FIGURE 90-5 Postgadolinium T1 MRI demonstrating multiple meningiomas along the falx and left parietal cortex.

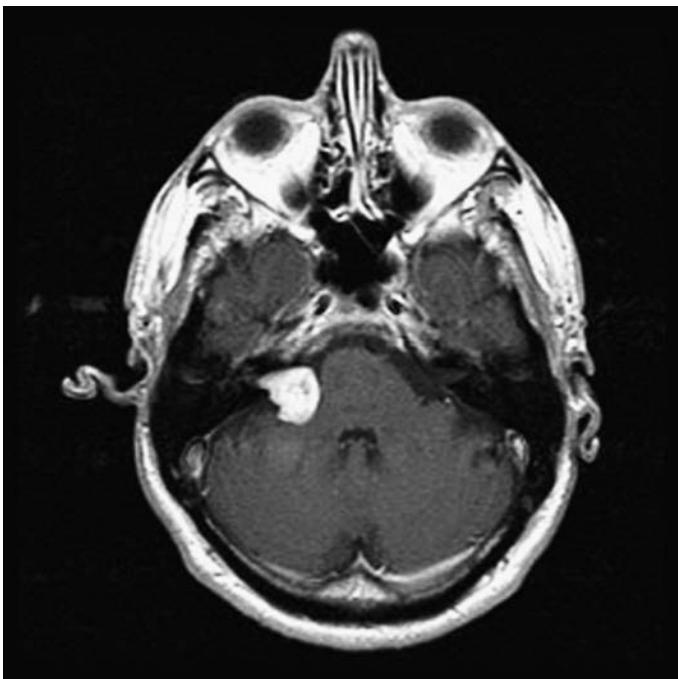


FIGURE 90-6 Postgadolinium MRI of a right vestibular schwannoma. The tumor can be seen to involve the internal auditory canal.

SCHWANNOMAS

These are generally benign tumors arising from the Schwann cells of cranial and spinal nerve roots. The most common schwannomas, termed *vestibular schwannomas* or *acoustic neuromas*, arise from the vestibular portion of the eighth cranial nerve and account for ~9% of primary brain tumors. Patients with NF2 have a high incidence of vestibular schwannomas that are frequently bilateral. Schwannomas arising from other cranial nerves, such as the trigeminal nerve (cranial nerve V), occur with much lower frequency. Neurofibromatosis type 1 (NF1) is associated with an increased incidence of schwannomas of the spinal nerve roots.

Vestibular schwannomas may be found incidentally on neuroimaging or present with progressive unilateral hearing loss, dizziness, tinnitus, or, less commonly, symptoms resulting from compression of the brainstem and cerebellum. On MRI, they appear as densely enhancing lesions, enlarging the internal auditory canal and often extending into the cerebellopontine angle (Fig. 90-6). The differential diagnosis includes meningioma. Very small, asymptomatic lesions can be observed with serial MRIs. Larger lesions should be treated with surgery or SRS. The optimal treatment will depend on the size of the tumor, symptoms, and the patient's preference. In patients with small vestibular schwannomas and relatively intact hearing, early surgical intervention increases the chance of preserving hearing.

PITUITARY TUMORS

These are discussed in detail in [Chap. 380](#).

CRANIOPHARYNGIOMAS

Craniopharyngiomas are rare, usually suprasellar, partially calcified, solid, or mixed solid-cystic benign tumors that arise from remnants of Rathke's pouch. They have a bimodal distribution, occurring predominantly in children but also between the ages of 55 and 65 years. They present with headaches, visual impairment, and impaired growth in children and hypopituitarism in adults. Treatment involves surgery, RT, or a combination of the two. The papillary subtype of craniopharyngiomas often has *BRAF* V600E mutations and can be treated with RAF/MEK inhibitors.

OTHER BENIGN TUMORS

Dysembryoplastic Neuroepithelial Tumors (DNTs) These are benign, supratentorial tumors, usually in the temporal lobe. They

typically occur in children and young adults with a long-standing history of seizures. Surgical resection is curative.

Epidermoid Cysts These consist of squamous epithelium surrounding a keratin-filled cyst. They are usually found in the cerebellopontine angle and the intrasellar and suprasellar regions. They may present with headaches, cranial nerve abnormalities, seizures, or hydrocephalus. MRI demonstrates an extra-axial lesion with characteristics that are similar to CSF but have restricted diffusion. Treatment involves surgical resection.

Dermoid Cysts Like epidermoid cysts, dermoid cysts arise from epithelial cells that are retained during closure of the neural tube. They contain both epidermal and dermal structures such as hair follicles, sweat glands, and sebaceous glands. Unlike epidermoid cysts, these tumors usually have a midline location. They occur most frequently in the posterior fossa, especially the vermis, fourth ventricle, and suprasellar cistern. On MRI, dermoid cysts resemble lipomas, demonstrating T1 hyperintensity and variable signal on T2. Symptomatic dermoid cysts can be treated with surgery.

Colloid Cysts These usually arise in the anterior third ventricle and may present with headaches, hydrocephalus, and, very rarely, sudden death. Surgical resection is curative, or a third ventriculostomy may relieve the obstructive hydrocephalus and be sufficient therapy.

NEUROCUTANEOUS SYNDROMES (PHAKOMATOSES)

A number of genetic disorders are characterized by cutaneous lesions and an increased risk of brain tumors. Most of these disorders have an autosomal dominant inheritance with variable penetrance.

NEUROFIBROMATOSIS TYPE 1 (von RECKLINGHAUSEN'S DISEASE)

NF1 is an autosomal dominant disorder with variable penetrance and an incidence of ~1 in 2600–3000. Approximately one-half of cases are familial; the remainder are caused by new mutations arising in patients with unaffected parents. The *NF1* gene is located on chromosome 17q11.2 and encodes neurofibromin, a guanosine triphosphatase (GTPase) activating protein (GAP) that is a negative regulator of the RAS-mitogen-activated protein (MAP) kinase signaling pathway, which includes the downstream kinase MEK. It is a classic tumor suppressor, and biallelic loss can result in a variety of nervous system tumors including neurofibromas, plexiform neurofibromas, optic nerve gliomas, astrocytomas, and meningiomas. In addition to neurofibromas, which appear as multiple, soft, rubbery cutaneous tumors, other cutaneous manifestations of NF1 include café-au-lait spots and axillary freckling. NF1 is also associated with hamartomas of the iris termed Lisch nodules, pheochromocytomas, pseudoarthrosis of the tibia, scoliosis, epilepsy, and mental retardation. The MEK inhibitor selumetinib has activity against inoperable plexiform neurofibromas and is the only treatment that targets the dysregulated signaling pathway.

NEUROFIBROMATOSIS TYPE 2

NF2 is less common than NF1, with an incidence of 1 in 25,000–40,000. It is an autosomal dominant disorder with full penetrance. As with NF1, approximately one-half of cases arise from new mutations. The *NF2* gene on 22q encodes a cytoskeletal protein, merlin (moesin, ezrin, radixin-like protein), that functions as a tumor suppressor. NF2 is characterized by bilateral vestibular schwannomas in >90% of patients, multiple meningiomas, and spinal ependymomas and astrocytomas. Treatment of bilateral vestibular schwannomas can be challenging because the goal is to preserve hearing for as long as possible. These patients may also have diffuse schwannomatosis that may affect the cranial, spinal, or peripheral nerves; posterior subcapsular lens opacities; and retinal hamartomas.

TUBEROUS SCLEROSIS (BOURNEVILLE DISEASE)

This is an autosomal dominant disorder with an incidence of ~1 in 5000–10,000 live births. It is caused by mutations in either the *TSC1*

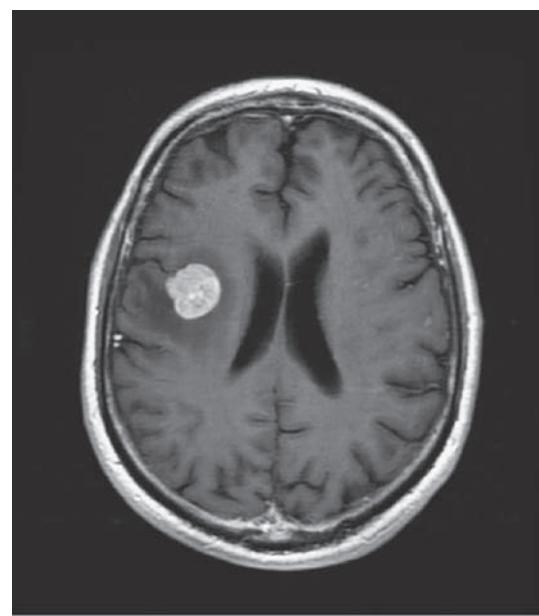
gene, which maps to chromosome 9q34 and encodes a protein termed hamartin, or the *TSC2* gene, which maps to chromosome 16p13.3 and encodes the protein tuberin. Hamartin forms a complex with tuberin, which inhibits cellular signaling through mTOR, and acts as a negative regulator of the cell cycle. Patients with tuberous sclerosis may have seizures, mental retardation, adenoma sebaceum (facial angiofibromas), shagreen patch, hypomelanotic macules, periungual fibromas, renal angiomyolipomas, and cardiac rhabdomyomas. These patients have an increased incidence of subependymal nodules, cortical tubers, and subependymal giant cell astrocytomas (SEGAs). Patients frequently require anticonvulsants for seizures. SEGAs do not always require therapeutic intervention, but the most effective therapy is with the mTOR inhibitors sirolimus or everolimus, which often decrease seizures as well as SEGA size.

TUMORS METASTATIC TO THE BRAIN

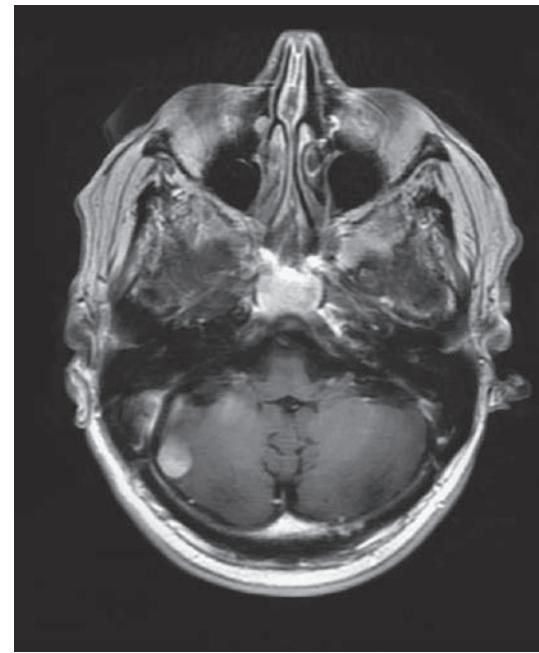
Brain metastases arise from hematogenous spread and frequently originate from a lung primary or are associated with pulmonary metastases. Most metastases develop at the gray matter–white matter junction in the watershed distribution of the brain where intravascular tumor cells lodge in terminal arterioles. The distribution of metastases in the brain approximates the proportion of blood flow such that ~85% of all metastases are supratentorial and 15% occur in the posterior fossa. The most common sources of brain metastases are lung and breast carcinomas; melanoma has the greatest propensity to metastasize to the brain, being found in 80% of patients at autopsy (Table 90-3). Other tumor types such as ovarian and esophageal carcinoma rarely metastasize to the brain. Prostate and breast cancers also have a propensity to metastasize to the dura and can mimic meningioma. Leptomeningeal metastases are common from hematologic malignancies and also breast and lung cancers. Spinal cord compression primarily arises in patients with prostate and breast cancer, tumors with a strong propensity to metastasize to the axial skeleton.

DIAGNOSIS OF METASTASES

Brain metastases are best visualized on MRI, where they usually appear as well-circumscribed lesions (Fig. 90-7). The amount of perilesional edema can be highly variable, with large lesions causing minimal edema and sometimes very small lesions causing extensive edema. Enhancement may be in a ring pattern or diffuse. Occasionally, intracranial metastases will hemorrhage; melanoma, thyroid, and kidney cancer have the greatest propensity to hemorrhage, but the most common cause of a hemorrhagic metastasis is lung cancer because it accounts for the majority of brain metastases. The radiographic appearance of brain metastasis is nonspecific, and similar-appearing lesions can occur with infection including brain abscesses, demyelinating lesions, sarcoidosis, radiation necrosis in a previously treated patient, or a primary brain tumor that may be a second malignancy in a patient with systemic cancer. Biopsy is rarely necessary for diagnosis because imaging alone in the appropriate clinical situation usually suffices. However, in ~10%



A



B

FIGURE 90-7 Postgadolinium T1 MRI of multiple brain metastases from non-small-cell lung cancer involving the right frontal (A) and right cerebellar (B) hemispheres. Note the diffuse enhancement pattern and absence of central necrosis.

TABLE 90-3 Frequency of Nervous System Metastases by Common Primary Tumors

	BRAIN (%)	LM (%)	ESCC (%)
Lung	41	17	15
Breast	19	57	22
Melanoma	10	12	4
Prostate	1	1	10
GIT	7	—	5
Renal	3	2	7
Lymphoma	<1	10	10
Sarcoma	7	1	9
Other	11	—	18

Abbreviations: ESCC, epidural spinal cord compression; GIT, gastrointestinal tract; LM, leptomeningeal metastases.

of patients, a systemic cancer may present with a brain metastasis, and if there is not an easily accessible systemic site to biopsy, a brain lesion must be removed for diagnostic purposes.

TREATMENT

Tumors Metastatic to the Brain

DEFINITIVE TREATMENT

The number and location of brain metastases often determine the therapeutic options. The patient's overall condition and current or potential control of systemic disease are also major determinants. Brain metastases are single in approximately one-half of patients and multiple in the other half.

RADIATION THERAPY

The standard treatment for brain metastases has previously been WBRT usually administered to a total dose of 3000 cGy in 10 fractions. This affords rapid palliation, and ~80% of patients improve with glucocorticoids and RT. However, it is not curative, is associated with neurocognitive toxicity, and produces median survival of only 4–6 months. Recent data demonstrate that hippocampal avoidance during WBRT preserves cognitive function without increasing the risk of an intracranial relapse. If feasible, SRS has become the primary radiation oncology approach to brain metastases. It can be delivered through a variety of equally effective techniques including the gamma knife, linear accelerator, proton beam, or CyberKnife, all of which can deliver highly focused doses of RT, usually in a single fraction. SRS can effectively sterilize the visible lesions and afford local disease control in 80–90% of patients. Some patients have been cured of their brain metastases using SRS, whereas this is distinctly rare with WBRT. Traditionally SRS was used only for patients with 1–3 metastases, but recent data suggest that SRS can effectively treat up to 10 lesions. It is, however, confined to lesions of ≤ 3 cm and is most effective in metastases of ≤ 1 cm. The addition of WBRT to SRS improves disease control in the nervous system but does not prolong survival and thus is rarely employed.

SURGERY

Randomized controlled trials have demonstrated that surgical extirpation of a single brain metastasis followed by WBRT is superior to WBRT alone. Removal of two lesions or a single symptomatic mass, particularly if compressing the ventricular system, can also be useful. This is particularly important in patients who have highly radioresistant lesions such as renal carcinoma. Surgical resection can produce rapid amelioration of symptoms, improve control of edema, and result in prolonged survival. WBRT administered after complete resection of a brain metastasis improves disease control but does not prolong survival. Some centers administer focal RT or even SRS to a resected cavity, especially if there is concern that tumor has been left behind, but most avoid postoperative WBRT because of its cognitive effects.

CHEMOTHERAPY

Chemotherapy is becoming increasingly useful for brain metastases. Metastases from tumor types that are highly chemosensitive, such as germ cell tumors or small-cell lung cancer, may respond to chemotherapeutic regimens chosen according to the underlying malignancy. Increasingly, data demonstrate responsiveness of brain metastases to chemotherapy including targeted therapeutics, such as for patients with lung cancer harboring EGFR mutations that sensitize them to EGFR inhibitors. Immunotherapy is also effective against those primary tumors that are sensitive to this approach, such as melanoma. Antiangiogenic agents such as bevacizumab are effective in the treatment of CNS metastases in those primary tumors for which it is approved.

LEPTOMENINGEAL METASTASES

Leptomeningeal metastases are also described as carcinomatous meningitis, meningeal carcinomatosis, or, in the case of specific tumors, leukemic or lymphomatous meningitis. Among the hematologic malignancies, acute leukemias most commonly metastasize to the subarachnoid space, followed in frequency by aggressive diffuse lymphomas. Among solid tumors, breast and lung carcinomas and melanoma most frequently spread in this fashion. Tumor cells reach the subarachnoid space via the arterial circulation or occasionally through retrograde flow in venous systems that drain metastases along the bony spine or cranium. In addition, leptomeningeal metastases may develop as a direct consequence of prior brain metastases and occur in almost 40% of patients who have a metastasis resected from the cerebellum.

CLINICAL FEATURES

Leptomeningeal metastases are characterized by multilevel symptoms and signs along the neuraxis. Combinations of lumbar and cervical

radiculopathies, cranial neuropathies, seizures, confusion, and encephalopathy from hydrocephalus or raised intracranial pressure can be present. Focal deficits such as hemiparesis or aphasia are rarely due to leptomeningeal metastases unless there is direct brain infiltration. New-onset limb pain in patients with breast cancer, lung cancer, or melanoma should prompt consideration of leptomeningeal spread.

LABORATORY AND IMAGING DIAGNOSIS

Leptomeningeal metastases are particularly challenging to diagnose because identification of tumor cells in the subarachnoid compartment may be elusive. MRI can be definitive when there are clear tumor nodules adherent to the cauda equina or spinal cord, enhancing cranial nerves, or subarachnoid enhancement on brain imaging (Fig. 90-8).



A



B

FIGURE 90-8 Postgadolinium MRI images of extensive leptomeningeal metastases from breast cancer. Nodules along the dorsal surface of the spinal cord (**A**) and cauda equina (**B**) are seen.

- 710** Imaging is diagnostic in ~75% of patients and is more often positive in patients with solid tumors. Demonstration of tumor cells in the CSF is definitive and often considered the gold standard. However, CSF cytologic examination is positive in only 50% of patients on the first lumbar puncture and still misses 10% after three CSF samples. New technologies, such as rare cell capture, enhance identification of tumor cells in the CSF; molecular profiling of the CSF can also identify tumor-specific mutations, indicating malignancy in the leptomeninges. CSF cytologic examination is most useful in hematologic malignancies, especially when combined with flow cytometry to identify a clonal population. Accompanying CSF abnormalities include an elevated protein concentration and an elevated white blood cell count; hypoglycorrachia is noted in <25% of patients but is useful when present. Identification of tumor markers may be helpful in some solid tumors.

TREATMENT

Leptomeningeal Metastases

The treatment of leptomeningeal metastasis is palliative because there is no curative therapy. RT to the symptomatically involved areas, such as skull base for cranial neuropathy, can relieve pain and sometimes improve function. Craniospinal irradiation (CSI) is avoided because it has significant toxicity with myelosuppression and gastrointestinal irritation as well as limited effectiveness. However, recent data on proton beam CSI suggest better disease control with fewer systemic toxicities. Systemic chemotherapy, targeted therapeutics, and immunotherapy have all demonstrated efficacy in the appropriate setting. Alternatively, intrathecal chemotherapy can be effective, particularly in hematologic malignancies. This is optimally delivered through an intraventricular cannula (Ommaya reservoir) rather than by lumbar puncture. Few drugs can be delivered safely into the subarachnoid space, and they have a limited spectrum of antitumor activity, perhaps accounting for the relatively poor response to this approach, particularly in solid tumors. In addition, impaired CSF flow dynamics can compromise intrathecal drug delivery. Surgery has a limited role in leptomeningeal metastasis. A ventriculoperitoneal shunt can relieve raised intracranial pressure; once placed, intrathecal drug cannot be used.

EPIDURAL METASTASIS

Epidural metastasis occurs in 3–5% of patients with a systemic malignancy and causes neurologic compromise by compressing the spinal cord or cauda equina. The most common cancers that metastasize to the epidural space are those malignancies that spread to bone, such as breast and prostate. Lymphoma can cause bone involvement and compression, but it can also invade an intervertebral foramen and cause spinal cord compression without bone destruction. The thoracic spine is affected most commonly, followed by the lumbar and then cervical spine.

CLINICAL FEATURES

Back pain is the presenting symptom of epidural metastasis in virtually all patients; the pain may precede neurologic findings by weeks or months. The pain is usually exacerbated by lying down; by contrast, arthritic pain is often relieved by recumbency. Leg weakness is seen in ~50% of patients, as is sensory dysfunction. Sphincter problems are present in ~25% of patients at diagnosis.

DIAGNOSIS

Diagnosis is established by imaging, preferably with an MRI of the entire spine (Fig. 90-9). Contrast is not required to identify bony or epidural lesions. Any patient with cancer who has severe back pain should undergo an MRI. Plain films, bone scans, or even CT scans may show bone metastases, but only MRI can reliably delineate epidural tumor. For patients unable to have an MRI, CT myelography should be performed to outline the epidural space. The differential diagnosis of epidural tumor includes epidural abscess, acute or chronic hematomas, epidural lipomatosis, and, rarely, extramedullary hematopoiesis.

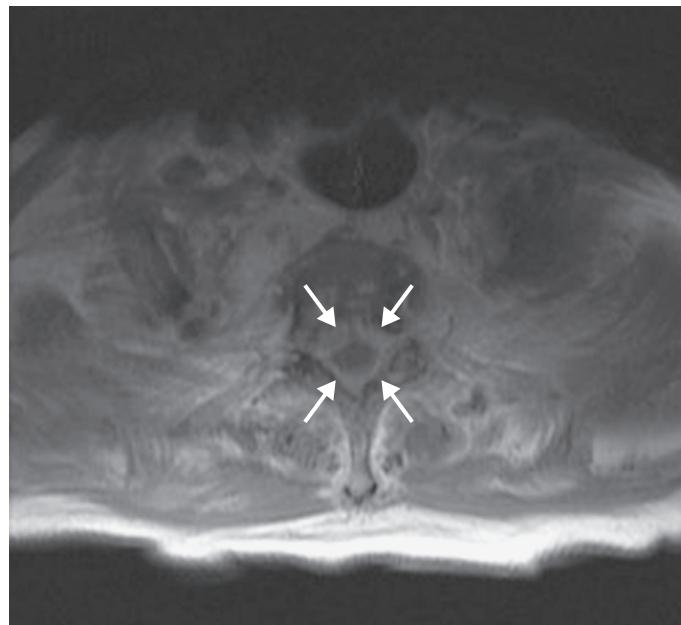


FIGURE 90-9 Postgadolinium T1 MRI showing circumferential epidural tumor around the thoracic spinal cord from esophageal cancer.

TREATMENT

Epidural Metastasis

Epidural metastasis requires immediate treatment. A randomized controlled trial demonstrated the superiority of surgical resection followed by RT compared to RT alone. However, patients must be able to tolerate surgery, and the surgical procedure of choice is a complete removal of the mass, which is typically anterior to the spinal canal, necessitating an extensive approach and resection. Otherwise, RT is the mainstay of treatment and can be used for patients with radiosensitive tumors, such as lymphoma, or for those unable to undergo surgery. SRS is increasingly being used, especially for radioresistant tumor types or for re-irradiation. Chemotherapy is rarely used for epidural metastasis unless the patient has minimal to no neurologic deficit and a highly chemosensitive tumor such as lymphoma or germinoma. Patients generally fare well if treated before there is a severe neurologic deficit. Recovery from paraparesis is better after surgery than with RT alone, but survival is often short due to widespread metastatic tumor.

NEUROLOGIC TOXICITY OF THERAPY

TOXICITY FROM RADIOTHERAPY

RT can cause a variety of toxicities in the CNS. These are usually described based on their relationship in time to the administration of RT: acute (occurring within days of RT), early delayed (months), or late delayed (years). In general, the acute and early delayed syndromes resolve and do not result in persistent deficits, whereas the late delayed toxicities are usually permanent and sometimes progressive.

Acute Toxicity Acute cerebral toxicity may occur during RT to the brain. RT can cause a transient disruption of the blood-brain barrier, resulting in edema and elevated intracranial pressure. This is usually manifest as headache, lethargy, nausea, and vomiting and can be both prevented and treated with the administration of glucocorticoids. There is no acute RT toxicity that affects the spinal cord.

Early Delayed Toxicity Early delayed toxicity is usually apparent weeks to months after completion of cranial irradiation and is likely due to focal demyelination. Clinically it may be asymptomatic or take the form of worsening or reappearance of a preexisting neurologic deficit. At times, a contrast-enhancing lesion can be seen on MRI/CT

that can mimic the tumor for which the patient received the RT. For patients with a malignant glioma, this has been described as “pseudoprogression” because it mimics tumor recurrence on MRI, but it represents inflammation and necrotic debris engendered by effective therapy. This is seen with increased frequency when chemotherapy, particularly temozolomide, is given concurrently with RT. Pseudoprogression can resolve on its own or, if very symptomatic, may require glucocorticoids, resection, or bevacizumab.

In the spinal cord, early delayed RT toxicity is manifest as a Lhermitte symptom with paresthesias of the limbs or along the spine when the patient flexes the neck. Although frightening, it is benign, resolves on its own, and does not portend more serious problems.

Late Delayed Toxicity Late delayed toxicities are the most serious because they are often irreversible and cause severe neurologic deficits. In the brain, late toxicities can take several forms, the most common of which include radiation necrosis and leukoencephalopathy. Radiation necrosis is a focal mass of necrotic tissue that is contrast enhancing on CT/MRI and may be associated with significant edema. This may appear identical to pseudoprogression but is seen months to years after RT and is always symptomatic. Clinical symptoms and signs include seizures and findings referable to the location of the necrotic mass. The necrosis is caused by the effect of RT on cerebral vasculature with fibrinoid necrosis and occlusion of blood vessels. It can mimic tumor radiographically, but unlike tumor, it is typically hypometabolic on a PET scan and has reduced perfusion on perfusion MR sequences. It may require resection for diagnosis and treatment unless it can be managed with glucocorticoids. There are reports of improvement with hyperbaric oxygen or bevacizumab, but symptomatic benefit does not always accompany radiographic improvement.

Leukoencephalopathy is seen most commonly after WBRT as opposed to focal RT. On T2 or FLAIR MR sequences, there is diffusely increased signal seen throughout the hemispheric white matter, often bilaterally and symmetrically. There tends to be a periventricular predominance that may be associated with atrophy and ventricular enlargement. Clinically, patients develop cognitive impairment, a gait disorder, and later urinary incontinence, all of which can progress over time. These symptoms mimic those of normal pressure hydrocephalus, and placement of a ventriculoperitoneal shunt can improve function in some patients but does not reverse the deficits completely. Increased age is a risk factor for leukoencephalopathy but not for radiation necrosis. Necrosis appears to depend on an unidentified predisposition.

Other late neurologic toxicities include endocrine dysfunction if the pituitary or hypothalamus was included in the RT port. An RT-induced neoplasm can occur many years after therapeutic RT for either a prior CNS or a head and neck tumor; accurate diagnosis requires surgical resection or biopsy. In addition, RT causes accelerated atherosclerosis, which can cause stroke either from intracranial vascular disease or carotid plaque from neck irradiation.

The peripheral nervous system is relatively resistant to RT toxicities. Peripheral nerves are rarely affected by RT, but the plexus is more vulnerable. Plexopathy develops more commonly in the brachial than in the lumbosacral distribution. It must be differentiated from tumor progression in the plexus, which is usually visualized by CT/MRI or PET scan demonstrating tumor infiltrating the region. Clinically, tumor progression is usually painful, whereas RT-induced plexopathy is painless. Radiation plexopathy is also more commonly associated with lymphedema and myokymia of the affected limb. Sensory loss and weakness are seen in both.

■ TOXICITY FROM CHEMOTHERAPY

Neurotoxicity is second to myelosuppression as the dose-limiting toxicity of chemotherapeutic agents (**Table 90-4**). Chemotherapy causes peripheral neuropathy from many commonly used agents, and the type of neuropathy can vary depending on the drug. Vincristine causes paresthesias but little sensory loss and is associated with motor dysfunction, autonomic impairment (frequently ileus), and, rarely, cranial nerve compromise. Cisplatin causes large-fiber sensory loss resulting in sensory ataxia but little cutaneous sensory loss and no weakness.

TABLE 90-4 Neurologic Signs Caused by Agents Commonly Used in Patients with Cancer

Acute encephalopathy (delirium)	Seizures
Methotrexate (high-dose IV, IT)	Methotrexate
Cisplatin	Etoposide (high-dose)
Vincristine	Cisplatin
Asparaginase	Vincristine
Procarbazine	Asparaginase
5-Fluorouracil (\pm levamisole)	Nitrogen mustard
Cytarabine (high-dose)	Carmustine
Nitrosoureas (high-dose or arterial)	Dacarbazine (intraarterial or high-dose)
Ifosfamide	Busulfan (high-dose)
Etoposide (high-dose)	Myelopathy (IT drugs)
Bevacizumab (PRES)	Methotrexate
Chronic encephalopathy (dementia)	Cytarabine
Methotrexate	Thiotepa
Carmustine	Peripheral neuropathy
Cytarabine	Vinca alkaloids
Fludarabine	Cisplatin
Visual loss	Procarbazine
Tamoxifen	Etoposide
Gallium nitrate	Teniposide
Cisplatin	Cytarabine
Fludarabine	Taxanes
Cerebellar dysfunction/ataxia	Suramin
5-Fluorouracil (\pm levamisole)	Bortezomib
Cytarabine	Procarbazine

Abbreviations: IT, intrathecal; IV, intravenous; PRES, posterior reversible encephalopathy syndrome.

The taxanes also cause a predominately sensory neuropathy. Agents such as bortezomib and thalidomide also cause neuropathy. Sometimes a severe neuropathy emerges after multiple neurotoxic agents have been used together or in sequence.

Encephalopathy and seizures are common toxicities from chemotherapeutic drugs. Ifosfamide can cause a severe encephalopathy, which is reversible with discontinuation of the drug and the use of methylene blue for severely affected patients. Fludarabine also causes a severe global encephalopathy that may be permanent. Bevacizumab and other anti-VEGF agents can cause posterior reversible encephalopathy syndrome. Cisplatin can cause hearing loss and less frequently vestibular dysfunction. Immunotherapy with monoclonal antibodies such as ipilimumab or nivolumab can cause an autoimmune hypophysitis, Guillain-Barré syndrome, or an autoimmune encephalitis.

■ FURTHER READING

- ACHROL AS et al: Brain metastases. *Nat Rev Dis Primers* 5:5, 2019.
- BARZILAI O et al: Predictors of quality of life improvement after surgery for metastatic tumors of the spine: Prospective cohort study. *Spine J* 18:1109, 2018.
- BROWN PD et al: Hippocampal avoidance during whole-brain radiotherapy plus memantine for patients with brain metastases: Phase III trial NRG oncology CC001. *J Clin Oncol* 38:1019, 2020.
- BUCKNER JC et al: Radiation plus procarbazine, CCNU, and vincristine in low-grade glioma. *N Engl J Med* 374:1344, 2016.
- CANCER GENOME ATLAS RESEARCH NETWORK et al: Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med* 372:2481, 2015.
- CHENG H, PEREZ-SOLER R: Leptomeningeal metastases in non-small-cell lung cancer. *Lancet Oncol* 19:e43, 2018.
- GROMMES C et al: Comprehensive approach to diagnosis and treatment of newly diagnosed primary CNS lymphoma. *Neuro Oncol* 21:296, 2019.
- GROSS AM et al: Selumetinib in children with inoperable plexiform neurofibromas. *N Engl J Med* 382:1430, 2020.

712 LOUIS DN et al: The 2016 World Health Organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol* 131:803, 2016.

McEWEN AE et al: Beyond the hood: CSF-derived cfDNA for diagnosis and characterization of CNS tumors. *Front Cell Dev Biol* 8:45, 2020.

MCGRANAHAN T et al: Current state of immunotherapy for treatment of glioblastoma. *Curr Treat Options Oncol* 20:24, 2019.

OMURO A et al: R-MVP followed by high-dose chemotherapy with TBC and autologous stem-cell transplant for newly diagnosed primary CNS lymphoma. *Blood* 125:1403, 2015.

RAMASWAMY V et al: Risk stratification of childhood medulloblastoma in the molecular era: The current consensus. *Acta Neuropathol* 131:821, 2016.

REARDON DA et al: Effect of nivolumab vs bevacizumab in patients with recurrent glioblastoma: The CheckMate 143 phase 3 randomized clinical trial. *JAMA Oncol* 6:1003, 2020.

RISHI A, YU HHM: Current treatment of melanoma brain metastasis. *Curr Treat Options Oncol* 21:45, 2020.

SCHIFF D et al: Recent developments and future directions in adult lower-grade gliomas: Society for Neuro-Oncology (SNO) and European Association of Neuro-Oncology (EANO) consensus. *Neuro Oncol* 21:837, 2019.

TSAKONAS G et al: Management of brain metastasized non-small cell lung cancer (NSCLC)—From local treatment to new systemic therapies. *Cancer Treat Rev* 54:122, 2017.

WEN PY et al: Glioblastoma in adults: A Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro Oncol* 22:1073, 2020.

YAMADA Y et al: The impact of histology and delivered dose on local control of spinal metastases treated with stereotactic radiosurgery. *Neurosurg Focus* 42:E6, 2017.

100,000 population, but the incidence varies with age. Soft tissue sarcomas constitute 0.7% of all cancers in the general population and 6.5% of all cancers in children.

■ EPIDEMIOLOGY

Malignant transformation of a benign soft tissue tumor is extremely rare, with the exception that malignant peripheral nerve sheath tumors (neurofibrosarcoma, malignant schwannoma) can arise from neurofibromas in patients with neurofibromatosis. Several etiologic factors have been implicated in soft tissue sarcomas.

Environmental Factors Trauma or previous injury is rarely involved, but sarcomas can arise in scar tissue resulting from a prior operation, burn, fracture, or foreign body implantation. Chemical carcinogens such as polycyclic hydrocarbons, asbestos, and dioxin may be involved in the pathogenesis.

Iatrogenic Factors Sarcomas in bone or soft tissues occur in patients who are treated with radiation therapy. The tumor nearly always arises in the irradiated field. The risk increases with time.

Viruses Kaposi's sarcoma (KS) in patients with HIV type 1, classic KS, and KS in HIV-negative homosexual men is caused by human herpesvirus (HHV) 8 ([Chap. 195](#)). No other sarcomas are associated with viruses.

Immunologic Factors Congenital or acquired immunodeficiency, including therapeutic immunosuppression, increases the risk of sarcoma.

■ GENETIC CONSIDERATIONS

 Li-Fraumeni syndrome is a familial cancer syndrome in which affected individuals have germline abnormalities of the tumor-suppressor gene *p53* and an increased incidence of soft tissue sarcomas and other malignancies, including breast cancer, osteosarcoma, brain tumor, leukemia, and adrenal carcinoma ([Chap. 71](#)). Neurofibromatosis 1 (NF-1, peripheral form, von Recklinghausen's disease) is characterized by multiple neurofibromas and café-au-lait spots. Neurofibromas occasionally undergo malignant degeneration to become malignant peripheral nerve sheath tumors. The gene for *NF1* is located in the pericentromeric region of chromosome 17 and encodes neurofibromin, a tumor-suppressor protein with guanosine 5'-triphosphate (GTP)ase-activating activity that inhibits ras function ([Chap. 90](#)). Germline mutation of the *RB1* locus (chromosome 13q14) in patients with inherited retinoblastoma is associated with the development of osteosarcoma in those who survive the retinoblastoma and of soft tissue sarcomas unrelated to radiation therapy. Other soft tissue tumors, including desmoid tumors, lipomas, leiomyomas, neuroblastomas, and paragangliomas, occasionally show a familial predisposition.

Ninety percent of synovial sarcomas contain a characteristic chromosomal translocation t(X;18)(p11;q11) involving a nuclear transcription factor on chromosome 18 called *SYT* and two breakpoints on X. Patients with translocations to the second X breakpoint (*SSX2*) may have longer survival than those with translocations involving *SSX1*.

Insulin-like growth factor (IGF) type II is produced by some sarcomas and may act as an autocrine growth factor and as a motility factor that promotes metastatic spread. IGF-II stimulates growth through IGF-I receptors, but its effects on motility are through different receptors. If secreted in large amounts, IGF-II may produce hypoglycemia ([Chaps. 93](#) and [406](#)). A large international sarcoma kindred study including 1162 patients and 6545 Caucasian controls revealed that about half the patients with sarcoma have putatively pathogenic mono-genic and polygenic variation in previously reported and new cancer genes, some of them representing therapeutically actionable targets. These patients were diagnosed with sarcoma at an earlier age compared to controls.

■ CLASSIFICATION

Approximately 20 different groups of sarcomas are recognized on the basis of the pattern of differentiation toward normal tissue. For example, rhabdomyosarcoma shows evidence of skeletal muscle fibers

91

Soft Tissue and Bone Sarcomas and Bone Metastases

Shreyaskumar R. Patel

Sarcomas are rare (<1% of all malignancies) mesenchymal neoplasms that arise in bone and soft tissues. These tumors are usually of mesodermal origin, although a few are derived from neuroectoderm, and they are biologically distinct from the more common epithelial malignancies. Sarcomas affect all age groups; 15% are found in children <15 years of age, and 40% occur after age 55 years. Sarcomas are one of the most common solid tumors of childhood and are the fifth most common cause of cancer deaths in children. Sarcomas may be divided into two groups, those derived from bone and those derived from soft tissues.

SOFT TISSUE SARCOMAS

Soft tissues include muscles, tendons, fat, fibrous tissue, synovial tissue, vessels, and nerves. Approximately 60% of soft tissue sarcomas arise in the extremities, with the lower extremities involved three times as often as the upper extremities. Thirty percent arise in the trunk, with the retroperitoneum accounting for 40% of all trunk lesions. The remaining 10% arise in the head and neck.

■ INCIDENCE

Approximately 13,130 new cases of soft tissue sarcomas occurred in the United States in 2020. The annual age-adjusted incidence is 3 per

with cross-striations; leiomyosarcomas contain interlacing fascicles of spindle cells resembling smooth muscle; and liposarcomas contain adipocytes. When precise characterization of the group is not possible, the tumors are called *unclassified sarcomas*. All of the primary bone sarcomas can also arise from soft tissues (e.g., extraskeletal osteosarcoma). The entity *malignant fibrous histiocytoma* (MFH) includes many tumors previously classified as fibrosarcomas or as pleomorphic variants of other sarcomas and is characterized by a mixture of spindle (fibrous) cells and round (histiocytic) cells arranged in a storiform pattern with frequent giant cells and areas of pleomorphism. As immunohistochemical suggestion of differentiation, particularly myogenic differentiation, may be found in a significant fraction of these patients, many are now characterized as poorly differentiated leiomyosarcomas, and the terms *undifferentiated pleomorphic sarcoma* (UPS) and *myxofibrosarcoma* are replacing MFH and myxoid MFH.

For purposes of treatment, most soft tissue sarcomas can be considered together. However, some specific tumors have distinct features. For example, *liposarcoma* can have a spectrum of behaviors. Pleomorphic liposarcomas and dedifferentiated liposarcomas behave like other high-grade sarcomas; in contrast, well-differentiated liposarcomas (better termed *atypical lipomatous tumors*) lack metastatic potential, and myxoid liposarcomas metastasize infrequently, but, when they do, they have a predilection for unusual metastatic sites containing fat, such as the retroperitoneum, mediastinum, and subcutaneous tissue. Rhabdomyosarcomas, Ewing's sarcoma, and other small-cell sarcomas tend to be more aggressive and are more responsive to chemotherapy than other soft tissue sarcomas.

Gastrointestinal stromal tumors (GISTs), previously classified as gastrointestinal leiomyosarcomas, are now recognized as a distinct entity within soft tissue sarcomas. Its cell of origin resembles the interstitial cell of Cajal, which controls peristalsis. The majority of malignant GISTs have activating mutations of the *c-kit* gene that result in ligand-independent phosphorylation and activation of the KIT receptor tyrosine kinase, leading to tumorigenesis. Approximately 5–10% of tumors will have a mutation in the platelet-derived growth factor receptor α (*PDGFRA*). GISTs that are wild type for both *KIT* and *PDGFRA* mutations may show mutations in *SDH B, C, or D* and may be driven by the IGF-I pathway.

■ DIAGNOSIS

The most common presentation is an asymptomatic mass. Mechanical symptoms referable to pressure, traction, or entrapment of nerves or muscles may be present. All new and persistent or growing masses should be biopsied, either by a small incision or by a cutting needle (core-needle biopsy) placed so that it can be encompassed in the subsequent excision without compromising a definitive resection. Lymph node metastases occur in 5%, except in synovial and epithelioid sarcomas, clear-cell sarcoma (melanoma of the soft parts), angiosarcoma, and rhabdomyosarcoma, where nodal spread may be seen in 17%. The pulmonary parenchyma is the most common site of metastases. Exceptions are GISTs, which metastasize to the liver; myxoid liposarcomas, which seek fatty tissue; and clear cell sarcomas, which may metastasize to bones. Central nervous system metastases are rare, except in alveolar soft part sarcoma.

Radiographic Evaluation Imaging of the primary tumor is best with plain radiographs and magnetic resonance imaging (MRI) for tumors of the extremities or head and neck and by computed tomography (CT) for tumors of the chest, abdomen, or retroperitoneal cavity. A radiograph and CT scan of the chest are important for the detection of lung metastases. Other imaging studies may be indicated, depending on the symptoms, signs, or histology.

■ STAGING AND PROGNOSIS

The histologic grade and size of the primary tumor are the most important prognostic factors. The current American Joint Committee on Cancer (AJCC) staging system is shown in Table 91-1. Prognosis is related to the stage. Cure is common in the absence of metastatic disease, but a small number of patients with metastases can also be cured. Historically, most patients with stage IV disease used to die within 12 months, but

TABLE 91-1 American Joint Commission on Cancer Staging System for Sarcomas, Eighth Edition

T1	Tumor ≤5 cm in greatest dimension
T2	Tumor >5 cm and ≤10 cm in greatest dimension
T3	Tumor >10 cm and ≤15 cm in greatest dimension
T4	Tumor >15 cm in greatest dimension
N0	No regional lymph node metastasis or unknown lymph node status
N1	Regional lymph node metastasis
M0	No distant metastasis
M1	Distant metastasis
Stage Groups	
Stage IA	T1; N0; M0; G1
Stage IB	T2, T3, T4; N0; M0; G1
Stage II	T1; N0; M0; G2/3
Stage IIIA	T1A, T2; N0; M0; G2/3
Stage IIIB	T3, T4; N0; M0; G2/3
Stage IV	Any T; N1; M0; any G Any T; any N; M1; any G

with availability of multiple lines of treatments, median survival in second-line and beyond ranges from 13 to 14 months, and some patients may live with stable or slowly progressive disease for many years.

TREATMENT

Soft Tissue Sarcomas

AJCC stage I patients are adequately treated with surgery alone. Stage II patients are considered for adjuvant radiation therapy. Stage III patients may benefit from neoadjuvant or adjuvant chemotherapy. Stage IV patients are managed primarily with systemic therapy, with or without other modalities.

SURGERY

Soft tissue sarcomas tend to grow along fascial planes, with the surrounding soft tissues compressed to form a pseudocapsule that gives the sarcoma the appearance of a well-encapsulated lesion. This is invariably deceptive because “shelling out,” or marginal excision, of such lesions results in a 50–90% probability of local recurrence. Wide excision with a negative margin, incorporating the biopsy site, is the standard surgical procedure for local disease. The adjuvant use of radiation therapy and/or chemotherapy improves the local control rate and permits the use of limb-sparing surgery with a local control rate (85–90%) comparable to that achieved by radical excisions and amputations. Limb-sparing approaches are indicated except when negative margins are not obtainable, when the risks of radiation are prohibitive, or when neurovascular structures are involved so that resection will result in serious functional consequences to the limb.

RADIATION THERAPY

External-beam radiation therapy is an adjuvant to limb-sparing surgery for improved local control. Preoperative radiation therapy allows the use of smaller fields and smaller doses but results in a higher rate of wound complications. Postoperative radiation therapy must be given to larger fields, because the entire surgical bed must be encompassed, and in higher doses to compensate for hypoxia in the operated field. This results in a higher rate of late complications. Brachytherapy or interstitial therapy, in which the radiation source is inserted into the tumor bed, is comparable in efficacy (except in low-grade lesions), less time consuming, and less expensive.

With the advent of stereotactic body radiotherapy (SBRT), the role of radiation therapy in oligometastatic disease in various visceral sites is being investigated and evolving.

ADJUVANT CHEMOTHERAPY

Chemotherapy is the mainstay of treatment for Ewing's sarcomas/primitive neuroectodermal tumors (PNETs) and rhabdomyosarcomas. Meta-analysis of 14 randomized trials in non-small-cell sarcomas revealed a significant improvement in local control and disease-free survival in favor of doxorubicin-based chemotherapy. Overall survival improvement was 4% for all sites and 7% for the extremity site. An updated meta-analysis including four additional trials with doxorubicin and ifosfamide combination reported a statistically significant 6% survival advantage in favor of chemotherapy. A chemotherapy regimen including an anthracycline and ifosfamide with growth factor support improved overall survival by 19% for high-risk (high-grade, ≥5 cm primary, or locally recurrent) extremity soft tissue sarcomas. Long-term follow-up of a trial evaluating neoadjuvant use of the same combination confirms survival advantage and reports a 10-year survival of 61%. A more contemporary randomized trial compared the standard anthracycline and ifosfamide combination to specific histology-tailored chemotherapy as an active control and confirmed superiority of the standard regimen.

ADVANCED DISEASE

Metastatic soft tissue sarcomas are largely incurable, but up to 20% of patients who achieve a complete response become long-term survivors. The therapeutic intent, therefore, is to produce a complete remission with chemotherapy (<10%) and/or surgery (30–40%). Surgical resection of metastases, whenever possible, is an integral part of the management. Some patients benefit from repeated surgical excision of metastases. The two most active chemotherapeutic agents are doxorubicin and ifosfamide. These drugs show a steep dose-response relationship in sarcomas. Gemcitabine with or without docetaxel has become an established second-line regimen and is particularly active in patients with UPS and leiomyosarcomas. Dacarbazine also has some modest activity. Taxanes have selective activity in angiosarcomas, and vincristine, etoposide, and irinotecan are effective in rhabdomyosarcomas and Ewing's sarcomas. Pazopanib, an inhibitor of the vascular endothelial growth factor, platelet-derived growth factor (PDGF), and c-kit, is now approved for patients with advanced soft tissue sarcomas excluding liposarcomas after failure of chemotherapy. Two additional chemotherapy drugs have gained approval from the U.S. Food and Drug Administration (FDA). Trabectedin was compared to dacarbazine in a large phase 3 randomized study in advanced leiomyosarcomas and liposarcomas after failure of an anthracycline and resulted in significant improvement in progression-free survival. Eribulin was also tested in a similar trial and showed improvement in survival, predominantly in the liposarcoma subgroup, and is therefore now approved for that subset. Tazemetostat, an EZH2 inhibitor, is now approved for use in metastatic epithelioid sarcomas characterized by loss of tumor-suppressor gene *INII*, resulting in activation of the EZH2 pathway. Imatinib targets KIT and PDGF tyrosine kinase activity and is standard therapy for advanced/metastatic GISTs and dermatofibrosarcoma protuberans. Imatinib is also indicated as adjuvant therapy for completely resected primary GISTs. Three years of adjuvant imatinib appear to be superior to 1 year of therapy for high-risk GISTs, although the optimal treatment duration remains unknown. Sunitinib and regorafenib are approved for second- and third-line use, respectively, in metastatic GIST after failure of or intolerance to imatinib. Ripretinib, an inhibitor of c-kit and PDGFRA, was approved for fourth-line use in metastatic GIST based on a placebo-controlled randomized trial reporting an improved median progression-free and overall survival. Avapritinib also received approval for use in the specific molecular subset of *PDGFRA* D842V-mutant metastatic GIST.

BONE SARCOMAS

INCIDENCE AND EPIDEMIOLOGY

Bone sarcomas are rarer than soft tissue sarcomas; they accounted for only 0.2% of all new malignancies and 3600 new cases in the United States in 2020. Several benign bone lesions have the potential for

malignant transformation. Enchondromas and osteochondromas can transform into chondrosarcoma; fibrous dysplasia, bone infarcts, and Paget's disease of bone can transform into either UPS or osteosarcoma.

CLASSIFICATION

Benign Tumors The common benign bone tumors include enchondroma, osteochondroma, chondroblastoma, and chondromyxoid fibroma, of cartilage origin; osteoid osteoma and osteoblastoma, of bone origin; fibroma and desmoplastic fibroma, of fibrous tissue origin; hemangioma, of vascular origin; and giant cell tumor, of unknown origin.

Malignant Tumors The most common malignant tumors of bone are plasma cell tumors ([Chap. 111](#)). The four most common malignant nonhematopoietic bone tumors are osteosarcoma, chondrosarcoma, Ewing's sarcoma, and UPS. Rare malignant tumors include chordoma (of notochordal origin), malignant giant cell tumor, adamantinoma (of unknown origin), and hemangioendothelioma (of vascular origin).

Musculoskeletal Tumor Society Staging System Sarcomas of bone are staged according to the Musculoskeletal Tumor Society staging system based on grade and compartmental localization. A Roman numeral reflects the tumor grade: stage I is low grade, stage II is high grade, and stage III includes tumors of any grade that have lymph node or distant metastases. In addition, the tumor is given a letter reflecting its compartmental localization. Tumors designated A are intracompartamental (i.e., confined to the same soft tissue compartment as the initial tumor), and tumors designated B are extracompartamental (i.e., extending into the adjacent soft tissue compartment or into bone). The tumor-node-metastasis (TNM) staging system is shown in [Table 91-2](#).

TABLE 91-2 Staging System for Bone Sarcomas

Primary tumor (T)	TX	Primary tumor cannot be assessed
	T0	No evidence of primary tumor
	T1	Tumor ≤8 cm in greatest dimension
	T2	Tumor >8 cm in greatest dimension
	T3	Discontinuous tumors in the primary bone site
Regional lymph nodes (N)	NX	Regional lymph nodes cannot be assessed
	N0	No regional lymph node metastasis
	N1	Regional lymph node metastasis
Distant metastasis (M)	MX	Distant metastasis cannot be assessed
	M0	No distant metastasis
	M1	Distant metastasis
	M1a	Lung
	M1b	Other distant sites
Histologic grade (G)	GX	Grade cannot be assessed
	G1	Well differentiated—low grade
	G2	Moderately differentiated—low grade
	G3	Poorly differentiated—high grade
	G4	Undifferentiated—high grade (Ewing's is always classed G4)

Stage Grouping

Stage IA	T1	N0	M0	G1,2 low grade
Stage IB	T2	N0	M0	G1,2 low grade
Stage IIA	T1	N0	M0	G3,4 high grade
Stage IIB	T2	N0	M0	G3,4 high grade
Stage III	T3	N0	M0	Any G
Stage IVA	Any T	N0	M1a	Any G
Stage IVB	Any T	N1	Any M	Any G
	Any T	Any N	M1b	Any G

■ OSTEOSARCOMA

Osteosarcoma, accounting for almost 45% of all bone sarcomas, is a spindle cell neoplasm that produces osteoid (unmineralized bone) or bone. Approximately 60% of all osteosarcomas occur in children and adolescents in the second decade of life, and ~10% occur in the third decade of life. Osteosarcomas in the fifth and sixth decades of life are frequently secondary to either radiation therapy or transformation in a preexisting benign condition, such as Paget's disease. Males are affected 1.5–2 times as often as females. Osteosarcoma has a predilection for metaphyses of long bones; the most common sites of involvement are the distal femur, proximal tibia, and proximal humerus. The classification of osteosarcoma is complex, but 75% of osteosarcomas fall into the "classic" category, which includes osteoblastic, chondroblastic, and fibroblastic osteosarcomas. The remaining 25% are classified as "variants" on the basis of (1) clinical characteristics, as in the case of osteosarcoma of the jaw, postradiation osteosarcoma, or Paget's osteosarcoma; (2) morphologic characteristics, as in the case of telangiectatic osteosarcoma, small-cell osteosarcoma, or epithelioid osteosarcoma; or (3) location, as in parosteal or periosteal osteosarcoma. Diagnosis usually requires a synthesis of clinical, radiologic, and pathologic features. Patients typically present with pain and swelling of the affected area. A plain radiograph reveals a destructive lesion with a moth-eaten appearance, a spiculated periosteal reaction (sunburst appearance), and a cuff of periosteal new bone formation at the margin of the soft tissue mass (Codman's triangle). A CT scan of the primary tumor is best for defining bone destruction and the pattern of calcification, whereas MRI is better for defining intramedullary and soft tissue extension. A chest radiograph and CT scan are used to detect lung metastases. Metastases to the bony skeleton should be imaged by a bone scan or by fluorodeoxyglucose positron emission tomography (FDG-PET). Almost all osteosarcomas are hypervascular and PET-avid. Pathologic diagnosis is established either with a core-needle biopsy, where feasible, or with an open biopsy with an appropriately placed incision that does not compromise future limb-sparing resection. Most osteosarcomas are high grade. The most important predictive factor for long-term survival is response to chemotherapy. Preoperative chemotherapy followed by limb-sparing surgery (which can be accomplished in >80% of patients) followed by postoperative chemotherapy is standard management. The effective drugs are doxorubicin, ifosfamide, cisplatin, and high-dose methotrexate with leucovorin rescue. The various combinations of these agents that have been used have all been about equally successful. Long-term survival rates in extremity osteosarcoma range from 60 to 80%. Osteosarcoma is radioresistant; radiation therapy has no role in the routine management. UPS is considered a part of the spectrum of osteosarcoma and is managed similarly. A randomized trial has shown improved progression-free survival with regorafenib compared to placebo.

■ CHONDROSARCOMA

Chondrosarcoma, which constitutes ~20–25% of all bone sarcomas, is a tumor of adulthood and old age with a peak incidence in the fourth to sixth decades of life. It has a predilection for the flat bones, especially the shoulder and pelvic girdles, but can also affect the diaphyseal portions of long bones. Chondrosarcomas can arise de novo or as a malignant transformation of an enchondroma or, rarely, of the cartilaginous cap of an osteochondroma. Chondrosarcomas have an indolent natural history and typically present as pain and swelling. Radiographically, the lesion may have a lobular appearance with mottled or punctate or annular calcification of the cartilaginous matrix. It is difficult to distinguish low-grade chondrosarcoma from benign lesions by x-ray or histologic examination. The diagnosis is therefore influenced by clinical history and physical examination. A new onset of pain, signs of inflammation, and progressive increase in the size of the mass suggest malignancy. The histologic classification is complex, but most tumors fall within the classic category. Like other bone sarcomas, high-grade chondrosarcomas spread to the lungs. Most chondrosarcomas are resistant to chemotherapy, and surgical resection of primary or recurrent tumors, including pulmonary metastases, is the mainstay of therapy with the exception of two histologic variants. Dedifferentiated chondrosarcoma has a high-grade osteosarcoma or a malignant fibrous histiocytoma component

that responds to chemotherapy. Mesenchymal chondrosarcoma, a rare variant composed of a small-cell element, also is responsive to systemic chemotherapy and is treated like Ewing's sarcoma.

■ EWING'S SARCOMA

Ewing's sarcoma, which constitutes ~10–15% of all bone sarcomas, is common in adolescence and has a peak incidence in the second decade of life. It typically involves the diaphyseal region of long bones and also has an affinity for flat bones. The plain radiograph may show a characteristic "onion peel" periosteal reaction with a generous soft tissue mass, which is better demonstrated by CT or MRI. This mass is composed of sheets of monotonous, small, round, blue cells and can be confused with lymphoma, embryonal rhabdomyosarcoma, and small-cell carcinoma. The presence of p30/32, the product of the *mic-2* gene (which maps to the pseudoautosomal region of the X and Y chromosomes), is a cell-surface marker for Ewing's sarcoma (and other members of the Ewing family of tumors, previously also called PNETs). Most PNETs arise in soft tissues; they include peripheral neuroepithelioma, Askin's tumor (chest wall), and esthesioneuroblastoma. Glycogen-filled cytoplasm detected by staining with periodic acid-Schiff is also characteristic of Ewing's sarcoma cells. The classic cytogenetic abnormality associated with this disease is a reciprocal translocation of the long arms of chromosomes 11 and 22, t(11;22), which creates a chimeric gene product of unknown function with components from the *fli-1* gene on chromosome 11 and *ews* on chromosome 22. This disease is very aggressive, and it is therefore considered a systemic disease. Common sites of metastases are lung, bones, and bone marrow. Systemic chemotherapy is the mainstay of therapy, often being used before surgery. Doxorubicin, cyclophosphamide or ifosfamide, etoposide, vincristine, and dactinomycin are active drugs. Topotecan or irinotecan in combination with an alkylating agent is often used in relapsed patients. Local treatment for the primary tumor includes surgical resection, usually with limb salvage or radiation therapy. Patients with lesions below the elbow and below the mid-calf have a 5-year survival rate of 80% with effective treatment. Ewing's sarcoma at first presentation is a curable tumor, even in the presence of obvious metastatic disease, especially in children <11 years old.

TUMORS METASTATIC TO BONE

Bone is a common site of metastasis for carcinomas of the prostate, breast, lung, kidney, bladder, and thyroid and for lymphomas and sarcomas. Prostate, breast, and lung primaries account for 80% of all bone metastases. Metastatic tumors of bone are more common than primary bone tumors. Tumors usually spread to bone hematogenously, but local invasion from soft tissue masses also occurs. In descending order of frequency, the sites most often involved are the vertebrae, proximal femur, pelvis, ribs, sternum, proximal humerus, and skull. Bone metastases may be asymptomatic or may produce pain, swelling, nerve root or spinal cord compression, pathologic fracture, or myelopathy (replacement of the marrow). Symptoms of hypercalcemia may be noted in cases of bony destruction.

Pain is the most frequent symptom. It usually develops gradually over weeks, is usually localized, and often is more severe at night. When patients with back pain develop neurologic signs or symptoms, emergency evaluation for spinal cord compression is indicated (Chap. 75). Bone metastases exert a major adverse effect on quality of life in cancer patients.

Cancer in the bone may produce osteolysis, osteogenesis, or both. Osteolytic lesions result when the tumor produces substances that can directly elicit bone resorption (vitamin D-like steroids, prostaglandins, or parathyroid hormone-related peptide) or cytokines that can induce the formation of osteoclasts (interleukin 1 and tumor necrosis factor). Osteoblastic lesions result when the tumor produces cytokines that activate osteoblasts. In general, purely osteolytic lesions are best detected by plain radiography, but they may not be apparent until they are >1 cm. These lesions are more commonly associated with hypercalcemia and with the excretion of hydroxyproline-containing peptides indicative of matrix destruction. When osteoblastic activity is prominent, the lesions may be readily detected using radionuclide

bone scanning (which is sensitive to new bone formation), and the radiographic appearance may show increased bone density or sclerosis. Osteoblastic lesions are associated with higher serum levels of alkaline phosphatase and, if extensive, may produce hypocalcemia. Although some tumors may produce mainly osteolytic lesions (e.g., kidney cancer) and others mainly osteoblastic lesions (e.g., prostate cancer), most metastatic lesions produce both types of lesion and may go through stages where one or the other predominates.

In older patients, particularly women, it may be necessary to distinguish metastatic disease of the spine from osteoporosis. In osteoporosis, the cortical bone may be preserved, whereas cortical bone destruction is usually noted with metastatic cancer.

TREATMENT

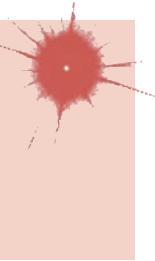
Metastatic Bone Disease

Treatment of metastatic bone disease depends on the underlying malignancy and the symptoms. Some metastatic bone tumors are curable (lymphoma, Hodgkin's disease), and others are treated with palliative intent. Pain may be relieved by local radiation therapy. Hormonally responsive tumors are responsive to hormone inhibition (antiandrogens for prostate cancer, antiestrogens for breast cancer). Strontium-89, samarium-153, and radium-223 are bone-seeking radionuclides that can exert antitumor effects and relieve symptoms. Denosumab, a monoclonal antibody that binds to RANK ligand, inhibits osteoclastic activity and increases bone mineral density. Bisphosphonates such as pamidronate may relieve pain and inhibit bone resorption, thereby maintaining bone mineral density and reducing risk of fractures in patients with osteolytic metastases from breast cancer and multiple myeloma. Careful monitoring of serum electrolytes and creatinine is recommended. Monthly administration prevents bone-related clinical events and may reduce the incidence of bone metastases in women with breast cancer. When the integrity of a weight-bearing bone is threatened by an expanding metastatic lesion that is refractory to radiation therapy, prophylactic internal fixation is indicated. Overall survival is related to the prognosis of the underlying tumor. Bone pain at the end of life is particularly common; an adequate pain relief regimen including sufficient amounts of narcotic analgesics is required. The management of hypercalcemia is discussed in **Chap. 410**.

FURTHER READING

- ALVAREZ RA et al: Optimization of the therapeutic approach to patients with sarcoma: Delphi consensus. *Sarcoma* 2019;4351308, 2019.
- BALLINGER ML et al: Monogenic and polygenic determinants of sarcoma risk: An international genetic study. *Lancet Oncol* 17:1261, 2016.
- DEMETRI GD et al: Efficacy and safety of Trabectedin or DTIC in patients with metastatic liposarcoma and leiomyosarcoma following failure of conventional chemotherapy: Results of a phase III randomized multicenter clinical trial. *J Clin Oncol* 34:786, 2016.
- GRONCHI A et al: Short, full-dose adjuvant chemotherapy (CT) in high-risk adult soft tissue sarcomas (STS): Long-term follow-up of a randomized clinical trial from the Italian Sarcoma Group and the Spanish Sarcoma Group. *Ann Oncol* 27:1, 2016.
- MEYER M, SEETHARAM M: First-line therapy for metastatic soft tissue sarcoma. *Curr Treat Options Oncol* 20:6, 2019.
- PASQUALI S, GRONCHI A: Neoadjuvant chemotherapy in soft tissue sarcomas: Latest evidence and clinical implications. *Ther Adv Med Oncol* 9:415, 2017.
- RATAN R, PATEL SR: Chemotherapy for soft tissue sarcoma. *Cancer* 122:2952, 2016.
- SCHOFFSKI P et al: Eribulin versus dacarbazine in previously treated patients with advanced liposarcoma or leiomyosarcoma: A randomised, open-label, multicentre, phase 3 trial. *Lancet* 387:1629, 2016.
- WAGNER MJ et al: Chemotherapy for bone sarcoma in adults. *J Oncol Pract* 12:208, 2016.

Kanwal Raghav, James L. Abbruzzese,
Gauri R. Varadhachary



Carcinoma (or cancer) of unknown primary (CUP) is a biopsy-proven malignancy for which the anatomic site of origin remains unidentified after a standardized detailed diagnostic evaluation. CUP is one of the 10 most frequently diagnosed cancers globally, accounting for 3–5% of all malignancies. Most investigators limit CUP to epithelial or undifferentiated cancers and do not include lymphomas, metastatic melanomas, and metastatic sarcomas because these cancers have specific histology and stage-based management guidelines, even in the absence of a primary site. CUP can occur in patients of all age groups including adolescents and young adults.

The emergence of sophisticated imaging, robust immunohistochemistry (IHC), and genomic and proteomic tools has challenged the “unknown” designation. Additionally, effective targeted therapies in several cancers and tissue agnostic biomarker-driven therapies have endorsed a change in paradigm from empiricism to a personalized approach to CUP management. The reasons cancers present as CUP remain unclear. One hypothesis is that the primary tumor either regresses after seeding the metastasis or remains so small that it is not detected. It is possible that CUP falls on the continuum of cancer presentation where the primary has been contained or eliminated by the natural body defenses, including the immune system. Alternatively, CUP may represent a specific malignant event that results in an increase in metastatic spread or survival relative to the primary. Whether the CUP metastases truly define a clone that is genetically and phenotypically unique to this diagnosis remains to be determined.

Since liver is a common site of CUP presentation, intrahepatic cholangiocarcinoma (ICC) can be often misdiagnosed as CUP. Of note, the incidence of ICC is increasing, whereas at the same time, that of CUP is declining. Improvements in diagnostic technologies including next-generation sequencing and other molecular techniques and awareness among clinicians to differentiate the two are possibly contributing to an increased recognition and incidence of ICC.

CUP BIOLOGY

Studies looking for unique signature abnormalities in CUP tumors have not been positive. Abnormalities in chromosomes 1 and 12 and other complex cytogenetic abnormalities have been reported. Aneuploidy has been described in 70% of CUP patients with metastatic adenocarcinoma or undifferentiated carcinoma. The overexpression of various genes, including *RAS*, *BCL2* (40%), *HER2* (11%), and *P53* (26–53%), has been identified in CUP samples, but they are found in many other malignancies and have no effect on response to therapy or survival. The extent of angiogenesis in CUP relative to that in metastases from known primaries has also been evaluated, but no consistent findings have emerged. Although current comprehensive genomic profiling efforts may help identify targeted therapeutic approaches to improve outcomes for this disease as discussed below, they have failed thus far to reveal a distinct molecular signature. More comprehensive and integrated multiomic efforts are needed to provide insights into CUP biology through recognition of molecular aberrations that especially drive metastatic growth.

APPROACH TO THE PATIENT

Carcinoma (or Cancer) of Unknown Primary

Initial CUP evaluation has two goals: search for the primary tumor based on pathologic evaluation of the metastases and determine the extent of disease. Focused evaluation directed by clinicopathologic

cues allows for judicious and efficient use of diagnostic tests. Obtaining a thorough medical history from CUP patients is essential, including paying particular attention to previous surgeries, removed lesions, and family medical history to assess potential hereditary cancers. Adequate physical examination, including a digital rectal examination in men and breast and pelvic examinations in women, should be performed based on clinical presentation. Finally, all patients with CUP, in the absence of contraindication, must undergo a computed tomographic (CT) scan of chest, abdomen and pelvis as a part of their standard work-up.

■ ROLE OF SERUM TUMOR MARKERS AND CYTOGENETICS

Most tumor markers, including carcinoembryonic antigen (CEA), CA-125, CA 19-9, and CA 15-3, when elevated, are nonspecific and not helpful in determining the primary site. Men who present with adenocarcinoma and predominant osteoblastic metastasis should undergo a prostate-specific antigen (PSA) test. In patients with undifferentiated or poorly differentiated carcinoma (especially with a midline tumor), elevated β -human chorionic gonadotropin (β -hCG) and α fetoprotein (AFP) levels suggest the possibility of an extragonadal germ cell (testicular) tumor. With the availability of advanced immunohistochemistry (IHC), cytogenetic studies are rarely needed.

■ ROLE OF IMAGING STUDIES

In the absence of contraindications, a baseline IV contrast computed tomography (CT) scan of the chest, abdomen, and pelvis is the standard of care. This helps to search for the primary tumor, evaluate the extent of disease, and select the most accessible biopsy site. With precise imaging and reporting, latent primary cancers, defined as appearance of a new primary cancer after a latent period of several months to years, is uncommon and seen in $\leq 5\%$ of CUP patients, usually in patients with very indolent presentations and/or highly responsive metastatic cancers that allows a latent primary to emerge (grow) over time.

Mammography should be performed in all women who present with metastatic adenocarcinoma, specifically in those with isolated axillary lymphadenopathy. Magnetic resonance imaging (MRI) of the breast can be considered in patients with axillary adenopathy and suspected occult primary breast carcinoma following a negative mammography and ultrasound. The results of these imaging modalities can influence surgical management; a negative MRI of the breast predicts a low tumor yield at mastectomy.

A conventional workup for a squamous cell carcinoma and cervical CUP (neck lymphadenopathy with no known primary tumor) includes a CT scan or MRI and invasive studies, including indirect and direct laryngoscopy, bronchoscopy, and upper endoscopy. Ipsilateral (or bilateral) staging tonsillectomy has been recommended for these patients. 18-Fluorodeoxyglucose positron emission tomography (18-FDG-PET) scans are useful in this patient population and may help guide the biopsy; determine the extent of disease; facilitate the appropriate treatment, including planning radiation fields; and help with disease surveillance. A smaller radiation field encompassing the metastatic adenopathy decreases the risk of chronic xerostomia. Several studies have evaluated the utility of PET in patients with squamous cervical CUP, and head and neck primary tumors were identified in $\sim 21\text{--}30\%$.

The diagnostic contribution of PET to the evaluation of other CUP presentations (outside of the neck adenopathy indication) remains controversial and is not routinely recommended. PET-CT can be helpful for patients with bone metastases and those deemed candidates for aggressive multimodality therapy (surgical intervention/radiation) such as patients with solitary metastatic disease because the identification of disease in addition to the solitary metastatic site may affect treatment planning.

Invasive studies, including upper endoscopy, colonoscopy, and bronchoscopy, should be limited to symptomatic patients or those with laboratory, imaging, or pathologic abnormalities that suggest that these techniques will result in a high yield in finding a primary cancer.

■ ROLE OF PATHOLOGIC STUDIES

A detailed pathologic examination of the most accessible biopsied tissue specimen is mandatory in CUP patients. Pathologic evaluation typically consists of hematoxylin and eosin stains and IHC tests. The importance of adequate tissue acquisition cannot be overemphasized in CUP. In addition to pathologic evaluation, tissue is also needed for tests of biomarkers of targeted agents, immunotherapy, and clinical trials.

Light Microscopy Evaluation Adequate tissue obtained preferably by excisional biopsy or core needle biopsy (instead of only a fine-needle aspiration) is stained with hematoxylin and eosin and subjected to light microscopic examination. On light microscopy, 60–65% of CUP is adenocarcinoma, and 5% is squamous cell carcinoma. The remaining 30–35% is poorly differentiated adenocarcinoma, poorly differentiated carcinoma, or poorly differentiated neoplasm. A small percentage of lesions are diagnosed as neuroendocrine cancers (2%), mixed tumors (adenosquamous or sarcomatoid carcinomas), or undifferentiated neoplasms (Table 92-1).

Role of IHC Analysis IHC stains are peroxidase-labeled antibodies against specific tumor antigens that are used to define tumor lineage. The number of available IHC stains is ever-increasing. However, a tiered and uniform approach to tissue evaluation in the CUP setting is lacking. For CUP cases, more is not necessarily better, and IHC stains should be used in conjunction with the patient's clinical presentation and imaging studies to select the best therapy. Communication between the clinician and pathologist is essential. No stain is 100% sensitive or specific, and under-/overinterpretation should be avoided. Poor differentiation, even in known primary tumors, decreases sensitivity of hallmark IHC markers. PSA and thyroglobulin tissue markers, which are positive in prostate and thyroid cancer, respectively, are the most specific of the current marker panel. However, these cancers rarely present as CUP, so the yield of these tests may be low. Figure 92-1 delineates a simple algorithm for immunohistochemical staining in CUP cases. Table 92-2 lists additional tests that may be useful to further define the tumor lineage. A more comprehensive algorithm may improve the diagnostic accuracy but can make the process complex and increase cost. With the use of IHC markers, electron microscopic analysis, which is time-consuming and expensive, is rarely needed.

There are >20 subtypes of cytokeratin (CK) intermediate filaments with different molecular weights and differential expression in various cell types and cancers. Monoclonal antibodies to specific CK subtypes have been used to help classify tumors according to their site of origin; commonly used CK stains in adenocarcinoma CUP are CK7 and CK20. CK7 is found in tumors of the lung, ovary, endometrium, breast, and upper gastrointestinal tract including pancreaticobiliary cancers, whereas CK20 is normally expressed in the gastrointestinal epithelium, urothelium, and Merkel cells. The nuclear CDX-2 transcription factor, which is the product of a homeobox gene necessary for intestinal organogenesis, is often used to aid in the diagnosis of gastrointestinal adenocarcinomas. However, CDX-2 positivity can be seen with enteric or mucinous differentiation in tumors from diverse primary sites (e.g., mucinous ovarian cancers).

Thyroid transcription factor 1 (TTF-1) nuclear staining is frequently positive in lung and thyroid cancers. Approximately 68% of adenocarcinomas and 25% of squamous cell lung cancers stain positive for TTF-1, which helps differentiate a lung primary tumor from metastatic

TABLE 92-1 Major Histologies in Carcinoma of Unknown Primary

HISTOLOGY	PROPORTION, %
Well to moderately differentiated adenocarcinoma	60
Squamous cell cancer	5
Poorly differentiated adenocarcinoma, poorly differentiated carcinoma	30
Neuroendocrine	2
Undifferentiated malignancy	3

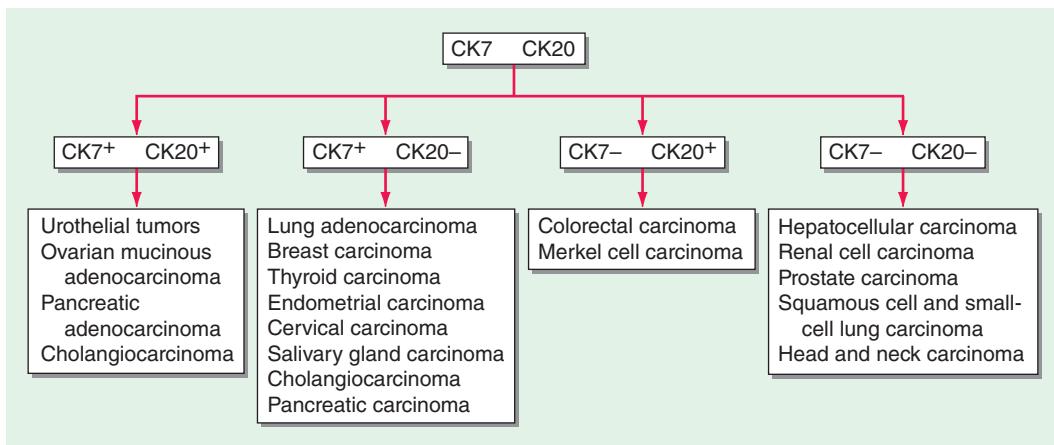


FIGURE 92-1 Approach to cytokeratin (CK7 and CK20) markers used in adenocarcinoma of unknown primary.

adenocarcinoma in a pleural effusion, the mediastinum, or the lung parenchyma.

Gross cystic disease fibrous protein-15 (GCDFP-15), a 15-kDa monomer protein, is a marker of apocrine differentiation that is detected in 62–72% of breast carcinomas. GATA3 is being increasingly used in the CUP setting when there is concern for a breast primary and can be particularly useful as a marker for metastatic breast carcinoma, especially triple-negative and metaplastic carcinomas, which lack specific endocrine markers of mammary origin. UROIII, high-molecular-weight cytokeratin, thrombomodulin, and CK20 are the markers used to diagnose lesions of urothelial origin.

TABLE 92-2 Select Immunohistochemical Stains Useful in the Diagnosis of CUP

LIKELY PRIMARY PROFILE	COMMONLY CONSIDERED IHC TO ASSIST IN DIFFERENTIAL DIAGNOSIS OF CUP ^a
Breast	ER, GCDFP-15, mammaglobin, HER2/neu, GATA3
Ovarian/mullerian	ER, WT1, CK7, PAX8, PAX2
Lung adenocarcinoma	TTF-1; nuclear staining, napsin A, SP-A1
Germ cell	β-hCG, AFP, OCT3/4, CKIT, CD30 (embryonal), SALL4
Prostate	PSA, α-methylacyl CoA racemase/P504S (AMACR/P504S), P501S (prostein), PSMA, NKX3-1
Intestinal	CK7, CK20, CDX-2, CEA
Neuroendocrine	Chromogranin, synaptophysin, CD56
Sarcoma	Desmin (desmoid tumors), factor VIII (angiosarcomas), CD31, smooth muscle actin (leiomyosarcoma), MyoD1 (rhabdomyosarcoma)
Renal	RCC, CD10, PAX8, CD10
Hepatocellular carcinoma	Hep Par-1, Arg-1, glypican-3
Melanoma	S100, SOX-10, vimentin, HMB-45, tyrosinase, melan-A
Urothelial	CK7, CK20, thrombomodulin, uroplakin III
Mesothelioma	Calretinin, WT1, D2-40, mesothelin
Lymphoma	LCA, CD3, CD4, CD5, CD20, CD45
SCC	p63, p40 (lung SCC), CK5/6

^aPatterns emerging from coexpression of stains are better than individual stains to suggest putative primary site. Even with optimization, no IHC panel is 100% sensitive or specific (e.g., ovarian mucinous carcinoma can exhibit positivity with intestinal markers).

Abbreviations: AFP, α fetoprotein; Arg-1, arginase-1; β-hCG, β-human chorionic gonadotropin; CEA, carcinoembryonic antigen; CUP, carcinoma of unknown primary; ER, estrogen receptor; GCDFP-15, gross cystic disease fibrous protein-15; IHC, immunohistochemistry; LCA, leukocyte common antigen; PSA, prostate-specific antigen; PSMA, prostate-specific membrane antigen; SCC, squamous cell carcinoma; SP-A1, surfactant protein A precursor; TTF, thyroid transcription factor; WT, Wilms' tumor.

IHC performs the best when used in groups that give rise to patterns that are strongly indicative of certain profiles. For example, the TTF-1/ CK7+ and CK20+/CDX-2+/CK7- phenotypes have been reported as very suggestive of lung and lower gastrointestinal cancer profiles, respectively. Despite their practical utility, these patterns have not been validated prospectively in CUP patients. IHC is not without its limitations; several factors affect tissue antigenicity (antigen retrieval, specimen processing, and fixation), interpretation of stains in tumor (nuclear, cytoplasmic, membrane) versus normal tissue, inter- and intraobserver variability, variable performance of different antibodies said to recognize the same antigen, and tissue heterogeneity and inadequacy (given small biopsy sizes). Communication with the pathologist is critical to determine if additional tissue will be beneficial in the pathologic evaluation. Pathologic features should not always supersede clinical or radiologic findings when considering testing for biomarkers of therapeutic response (e.g., epidermal growth factor receptor [EGFR], ALK mutations, human epidermal growth factor receptor 2 [HER-2]).

Role of Cancer Classifier Molecular Profiling In the absence of a known primary, developing therapeutic strategies for CUP is challenging. The current diagnostic yield with imaging and immunohistochemistry is ~20–30% for CUP patients. To reduce diagnostic uncertainty, sophisticated molecular analytics have been applied to CUP samples. These include gene expression profiling, messenger RNA (mRNA), microRNA, and epigenetic profiling to classify the CUP cancer.

Gene expression profiles are most commonly generated using quantitative reverse transcriptase polymerase chain reaction (RT-PCR) or DNA microarray. Neural network programs are then used to develop predictive algorithms from the gene expression profiles. Typically, a training set of gene profiles from known cancers (preferably from metastatic sites) is used to train the software. Comprehensive gene expression databases that have become available for common malignancies are then applied to CUP samples, and the program can then be used to predict the putative origin of a CUP sample.

mRNA- or microRNA-based tissue of origin cancer classifier assays have also been studied in prospective and retrospective CUP trials. More recently, a classifier based on microarray DNA methylation signatures has been studied and validated in known cancers. The DNA methylation profiling predicted a primary cancer in 87% of the 216 CUP patients.

Despite the sophistication of the cancer classifier molecular assays, most of the CUP studies have evaluated assay *performance*, although the challenge with validating the accuracy of an assay for CUP is that, by definition, the primary cancer diagnosis cannot be verified. Thus, current estimates of tissue of origin test accuracy have relied on indirect metrics, including comparison with pathology/IHC, clinical presentation, appearance of latent primaries, and autopsies. Using

these measures, the assays suggest a plausible primary in ~70–80% of patients studied. Three outcomes-based studies have been performed. First, a single-arm study reported a median survival of 12.5 months for patients who received assay-directed site-specific therapy. Second, a phase 2 trial of site-specific therapy, including molecularly targeted therapy, based on predicted tumor site from an algorithm using gene expression and alteration profile showed a 1-year survival of 53.1%. However, a randomized clinical trial evaluating site-specific therapy directed by gene expression profiling versus empirical chemotherapy with paclitaxel and carboplatin failed to show a significant improvement in 1-year survival (44% vs 55%, $p = .264$) with this approach. Firm conclusions of therapeutic impact cannot be drawn from these studies given the sample size, design, statistical biases, confounding variables including use of subsequent lines of (empiric) therapy, and heterogeneity of the CUP cancers. Additional studies are needed to better understand the clinical impact of tissue of origin profiling tools and how these assays complement IHC and help guide therapy.

Role of Next-Generation Sequencing A significant push is being made toward personalized medicine across all cancer types with the goal of identifying driver mutation(s) in a patient who can be treated with targeted agents independent of the site of origin. A retrospective study of 200 CUP tumor specimens reported on genomic alterations using the hybrid capture-based FoundationOne assay. The authors reported that a large number of CUP samples (85%) harbored at least one clinically relevant genomic alteration with the potential to influence and personalize therapy. The mean number of genomic alterations was 4.2 per tumor, and the most common genetic alterations included *TP53* (55%), *KRAS* (20%), *CDKN2A* (19%), and *ARID1A* (11%). The adenocarcinoma CUP tumors were more frequently driven by genetic alterations in the receptor tyrosine kinase (RTK)/Ras/mitogen-activated protein kinase (MAPK) signaling pathway than nonadenocarcinoma CUP tumors. Although, druggable genetic lesions seen in CUP are comparable to those in defined large entities, whether molecularly stratified approaches for CUP will successfully improve outcomes remains to be seen and clinical trials are needed. In a single-arm phase 2 study of 97 patients with molecularly

targeted therapy, five patients were found to have targetable *EGFR* mutations. Of these, four patients were treated with afatinib, an anti-*EGFR* drug, and two patients achieved a progression-free survival of >6 months. The emerging role of assays looking for circulating tumor cells, so-called liquid biopsies, within known tumor types has stirred interest in their potential utility in CUP.

Ongoing histology and cellular-context agnostic prospective clinical trials are studying the presence of actionable mutations and matching patients to the right targeted drug. Should this approach eventually be appropriately validated, CUP would be a natural fit for genomic alteration (GA)-based targeted therapy independent of tumor site. Immune checkpoint inhibitors (pembrolizumab) for microsatellite instability high (MSI-H) or deficient mismatch repair (dMMR) tumors and NTRK inhibitors for *NTRK* fusion-positive tumors can help a small minority of CUP patients.

TREATMENT

Carcinoma (or Cancer) of Unknown Primary

GENERAL CONSIDERATIONS

The treatment of CUP continues to evolve, albeit slowly. The median survival of most patients with disseminated CUP is ~6–10 months. Systemic chemotherapy is the primary treatment modality in most patients with disseminated disease, but the careful integration of surgery, radiation therapy, and even periods of observation is important in the overall management of this condition (Figs. 92-2 and 92-3). Prognostic factors include performance status, site and number of metastases, response to chemotherapy, and serum lactate dehydrogenase (LDH) levels. Culine and colleagues developed a prognostic model using performance status and serum LDH levels, which allowed the assignment of patients into two subgroups with divergent outcomes. Raghav and colleagues developed a prognostic nomogram to provide individualized survival estimates for patients with CUP based on baseline gender, ECOG performance status, histology, number of metastatic

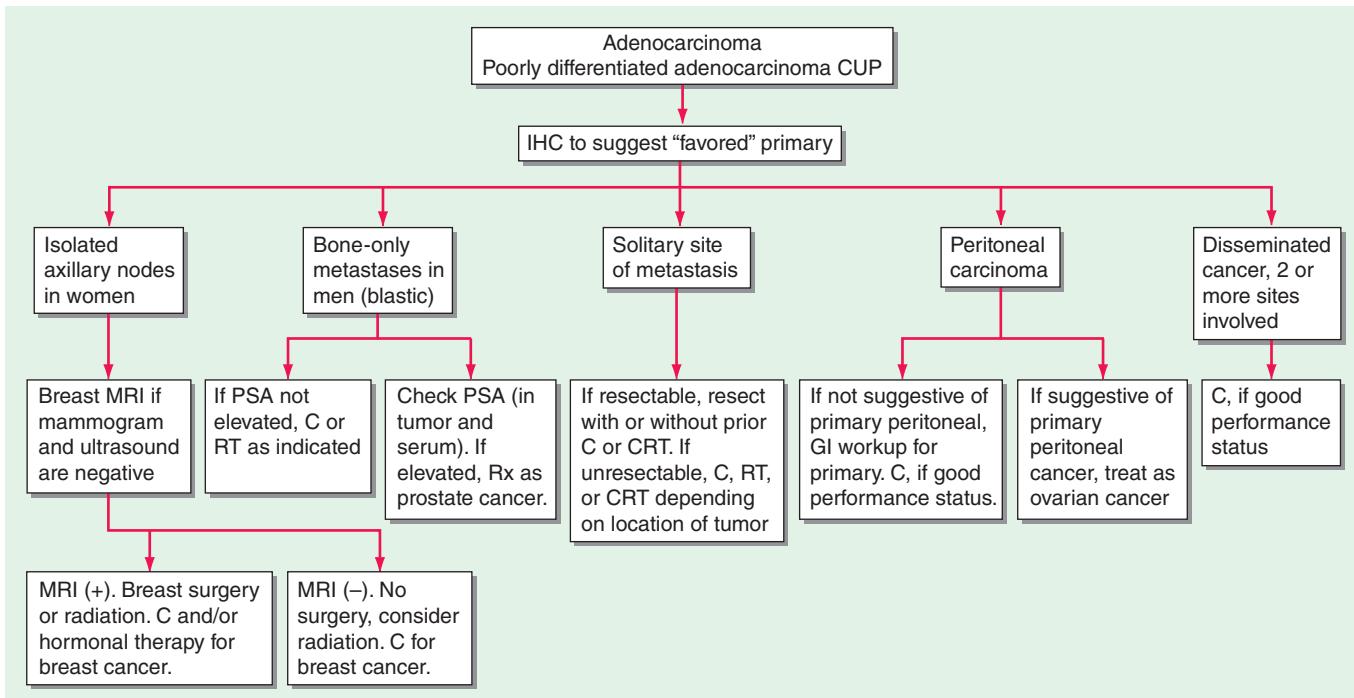


FIGURE 92-2 Treatment algorithm for adenocarcinoma and poorly differentiated adenocarcinoma of unknown primary (CUP). C, chemotherapy; CRT, chemoradiation; GI, gastrointestinal; IHC, immunohistochemistry; MRI, magnetic resonance imaging; PSA, prostate-specific antigen; RT, radiation.

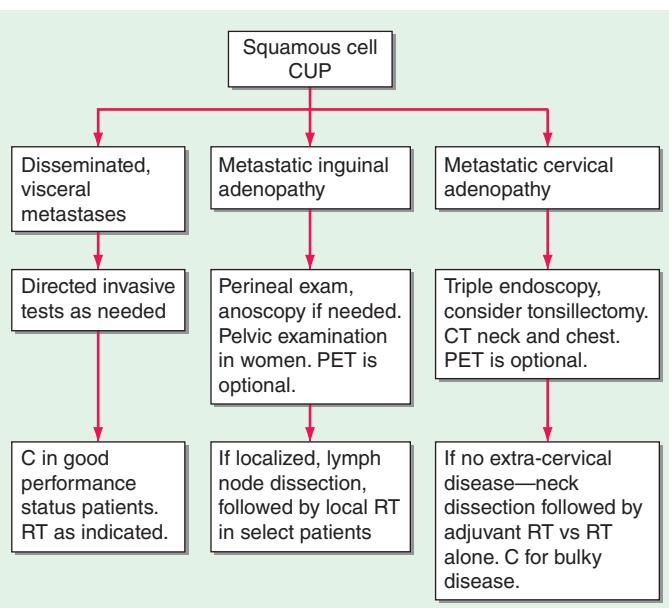


FIGURE 92-3 Treatment algorithm for squamous cell carcinoma of unknown primary (CUP). C, chemotherapy; CT, computed tomography; PET, positron emission tomography; RT, radiation.

sites and neutrophil-lymphocyte ratio. Future prospective trials using this prognostic model are warranted. Clinically, some CUP diagnoses fall into a favorable prognostic subset. Others, including those with disseminated CUP, have a more unfavorable prognosis.

TREATMENT OF FAVORABLE CUP SUBSETS

Women with Isolated Axillary Adenopathy Women with isolated axillary adenopathy with adenocarcinoma or carcinoma are usually treated for stage II or III breast cancer based on pathologic findings. These patients should undergo a breast MRI if mammogram and ultrasound are negative. Radiation therapy to the ipsilateral breast is indicated if the MRI of the breast is positive. Chemotherapy and/or hormonal therapy are indicated based on patient's age (premenopausal or postmenopausal), nodal disease bulk, and hormone receptor and HER2 status (Chap. 79). It is important to verify that the pathology suggests a breast cancer profile (morphology, IHC breast markers including estrogen receptor, mammaglobin, GCDFP-15, GATA3, HER-2 gene expression) before embarking on a breast cancer therapeutic program.

Women with Peritoneal Carcinomatosis The term *primary peritoneal papillary serous carcinoma* (PPSC) has been used to describe CUP with carcinomatosis with the pathologic and laboratory (elevated CA-125 antigen) characteristics of ovarian cancer but no ovarian primary tumor identified on transvaginal sonography or laparotomy. Studies suggest that ovarian cancer and PPSC, which are both of müllerian origin, have similar gene expression profiles. Similar to patients with ovarian cancer, patients with PPSC are candidates for cytoreductive surgery, followed by adjuvant taxane- and platinum-based chemotherapy. In one retrospective study of 258 women with peritoneal carcinomatosis who had undergone cytoreductive surgery and chemotherapy, 22% of patients had a complete response to chemotherapy; the median survival duration was 18 months (range 11–24 months). However, not all peritoneal carcinomatosis in women is PPSC. Careful pathologic evaluation can help diagnose a colon cancer profile (CDX-2+, CK20+, CK7-) or a pancreaticobiliary cancer or even a mislabeled peritoneal mesothelioma (calretinin, D2-40 positive; BerEP4, MOC-31 negative).

Poorly Differentiated Carcinoma with Midline Adenopathy (Chap. 88) Men with poorly differentiated or undifferentiated carcinoma who present with midline adenopathy should be evaluated for extragonadal germ cell malignancy. If diagnosed and treated as such, they often experience a good response to treatment with platinum-based combination chemotherapy. Response rates of >50% have been noted, and long-term survival rates of 10–15% have been reported. Older patients, especially smokers, who present with mediastinal adenopathy are more likely to have a lung or head and neck cancer profile.

Neuroendocrine Cancer (Chap. 84) Low-grade neuroendocrine tumor (NET) often has an indolent course, and treatment decisions are based on symptoms and tumor bulk. Urine 5-HIAA and serum chromogranin may be elevated and can be followed as markers. Often the patient is treated with somatostatin analogues alone for hormone-related symptoms (diarrhea, flushing, nausea). Specific local therapies or systemic therapy would only be indicated if the patient is symptomatic with local pain secondary to significant growth of the metastasis or the hormone-related symptoms are not controlled with endocrine therapy. Novel therapy options have demonstrated benefit in patients with low-grade NET, including sunitinib (which targets the vascular endothelial growth factor pathway), everolimus (which inhibits the mammalian target of rapamycin), and lutetium-177 dotatate (a somatostatin peptide receptor radioligand). Patients with high-grade NET are treated with platinum-based doublet therapy; 20–25% show a complete response, and up to 10% patients with limited/oligo presentations survive for >5 years. Some degree of neuroendocrine differentiation can be seen in diverse poorly differentiated carcinomas.

Squamous Cell Carcinoma Presenting as Neck Adenopathy Patients with early-stage squamous cell carcinoma involving the cervical lymph nodes are candidates for node dissection and radiation therapy, which can result in long-term survival. The role of chemotherapy in these patients is undefined, although chemoradiation therapy or induction chemotherapy is often used and is beneficial in bulky N2/N3 lymph node disease.

Solitary Metastatic Sites Patients with solitary metastases can also experience good treatment outcomes. Some patients who present with locoregional disease are candidates for aggressive trimodality (chemotherapy, radiation, and surgery) management—both prolonged disease-free survival and, occasionally, cure are possible.

Men with Blastic Skeletal Metastases and Elevated PSA (Chap. 87) Blastic bone-only metastasis is a rare presentation, and elevated serum PSA or tumor staining with PSA may provide confirmatory evidence of prostate cancer in these patients. Those with elevated levels are candidates for hormonal or other therapy for prostate cancer, although it is important to rule out other primary tumors (lung most common).

MANAGEMENT OF DISSEMINATED CUP

Patients who present with liver, brain, and adrenal metastatic disease usually have a poor prognosis. Patients with peritoneal carcinomatosis secondary to metastatic adenocarcinoma have a broad differential diagnosis, which includes mainly gastrointestinal cancers including gastric, appendiceal, colon, and pancreaticobiliary cancers.

Traditionally, platinum-based combination chemotherapy regimens have been used to treat CUP. Several broadly used regimens have been studied in the past two decades; these include paclitaxel-carboplatin, gemcitabine-cisplatin, gemcitabine-oxaliplatin, and irinotecan and fluoropyrimidine-based therapies. These chemotherapeutic agents used as empiric regimens have shown response rates of 25–40%, and their use obtains median survival times of 6–13 months.

J. Larry Jameson, Dan L. Longo

Outside of favorable subsets, there is a small group of patients with a “definitive” IHC profile. These patients usually have a single diagnosis based on their clinicopathologic presentation and are often treated for the putative primary tumor. This does not guarantee a response, although it increases the probability of response when select drugs are chosen from a class of drugs known to be effective in that cancer type. Efforts should be made to search for biomarkers of response to tumor-agnostic effective therapies such as immunotherapy for MSI-H/dMMR tumors. Patients who do not fall into those categories are candidates for broad-spectrum platinum-based regimens, clinical trials, and additional trial-based genomic and proteomic tests. Today, we do not have many effective drugs for several CUP cancer profiles, and treatments overlap for some cancers. Immunotherapy has been an area of active interest due to robust and durable responses in cancers with known primaries and has shown some activity in CUP. However, biomarkers of response and immune-sensitive subsets need to be defined within CUP.

SUMMARY

Patients with CUP should undergo a directed diagnostic search for the primary tumor on the basis of clinical and pathologic data. Subsets of patients have prognostically favorable disease, as defined by clinical or histologic criteria, and may substantially benefit from aggressive treatment; in these patients, prolonged survival can be expected. However, for most patients who present with advanced CUP, the prognosis remains poor with early resistance to available cytotoxic therapy. The current focus has shifted away from empirical chemotherapeutic trials to understanding the metastatic phenotype, tissue of origin profiling in select patients, and next-generation sequencing to identify actionable mutations in CUP patients. As novel therapies evolve in cancers with known primaries, investigations to assess their value in CUP will likely have an impact on management of CUP patients.

FURTHER READING

- FIZAZI K et al: Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 26(Suppl 5):v133, 2015.
- GATALICA Z et al: Comprehensive analysis of cancers of unknown primary for the biomarkers of response to immune checkpoint blockade therapy. Eur J Cancer 94:179, 2018.
- HAYASHI H et al: Randomized phase II trial comparing site-specific treatment based on gene expression profiling with carboplatin and paclitaxel for patients with cancer of unknown primary site. J Clin Oncol 37:570, 2019.
- HAYASHI H et al: Site-specific and targeted therapy based on molecular profiling by next-generation sequencing for cancer of unknown primary site: A nonrandomized phase 2 clinical trial. JAMA Oncol 6:1, 2020.
- NATIONAL COMPREHENSIVE CANCER NETWORK: Occult Primary (Cancer of Unknown Primary) version 2.2017, October 2016. http://www.nccn.org/professionals/physician_gls/pdf/occult.pdf.
- RAGHAV K et al: Cancer of unknown primary in adolescents and young adults: Clinicopathological features, prognostic factors and survival outcomes. PLoS One 11:e0154985, 2016.
- RAGHAV K et al: Defining a distinct immunotherapy eligible subset of patients with cancer of unknown primary using gene expression profiling with the 92-gene assay. Oncologist 25:e1807, 2020.
- RAGHAV K et al: Development and validation of a novel nomogram for individualized prediction of survival in cancer of unknown primary. Clin Cancer Res 27:3414, 2021.
- ROSS JS et al: Comprehensive genomic profiling of carcinoma of unknown primary site: New routes to targeted therapies. JAMA Oncol 1:40, 2015.
- VARADHACHARY GR, RABER MN: Cancer of unknown primary site. N Engl J Med 371:757, 2014.

Neoplastic cells can produce a variety of substances that can alter the physiology of hormonal, hematologic, dermatologic, rheumatologic, renal, and neurologic systems. *Paraneoplastic syndromes* refer to the disorders that accompany benign or malignant tumors but are not directly related to mass effects or invasion. Tumors of neuroendocrine origin, such as small-cell lung carcinoma (SCLC) and carcinoids are common causes of paraneoplastic syndromes, but they have been associated with many types of tumors that produce peptide hormones, cytokines, and growth factors and induce the production of antibodies. Studies of the prevalence of paraneoplastic syndromes indicate that they are more common than is generally appreciated. The signs, symptoms, and metabolic alterations associated with paraneoplastic disorders are easily overlooked in the context of a malignancy and its treatment. Consequently, atypical clinical manifestations in a patient with cancer should prompt consideration of a paraneoplastic syndrome. The most common hormonal and hematologic syndromes associated with underlying neoplasia will be discussed here.

ENDOCRINE PARANEOPLASTIC SYNDROMES

Etiology Hormones can be produced from eutopic or ectopic sources. *Eutopic* refers to the expression of a hormone from its normal tissue of origin, whereas *ectopic* refers to hormone production from an atypical tissue source. For example, adrenocorticotrophic hormone (ACTH) is expressed eutopically by the corticotrope cells of the anterior pituitary, but it can be expressed ectopically in SCLC. Many hormones are produced at low levels from tissues other than the classic endocrine source. Thus, ectopic expression is often a quantitative change rather than an absolute change in tissue expression. Nevertheless, the term *ectopic expression* is firmly entrenched and conveys the abnormal physiology associated with hormone production by neoplastic cells. In addition to high levels of hormones, ectopic expression is often characterized by abnormal regulation of hormone production (e.g., defective feedback control in ectopic ACTH) and peptide processing (resulting in large, unprocessed precursor peptide such as proopiomelanocortin [POMC]).

Many different molecular mechanisms can cause ectopic hormone production. In rare instances, genetic rearrangements account for aberrant hormone expression. For example, translocation of the parathyroid hormone (*PTH*) gene can result in high levels of *PTH* expression in tissues other than the parathyroid gland because the genetic rearrangement brings the *PTH* gene under the control of atypical regulatory elements. A related phenomenon is well documented in many forms of leukemia and lymphoma, in which somatic genetic rearrangements confer a growth advantage and alter cellular differentiation and function. Although genetic rearrangements cause selected cases of ectopic hormone production, this mechanism is rare, as many tumors are associated with excessive production of numerous peptides. Cellular dedifferentiation probably underlies most cases of ectopic hormone production. Many cancers are poorly differentiated, and certain tumor products, such as human chorionic gonadotropin (hCG), *PTH*-related protein (*PTHRP*), and a fetoprotein, are characteristic of gene expression at earlier developmental stages. In contrast, the propensity of certain cancers to produce particular hormones (e.g., squamous cell carcinomas produce *PTHRP*) suggests that dedifferentiation is partial or that selective pathways are derepressed. These expression profiles probably reflect epigenetic modifications that alter transcriptional

repression, microRNA expression, and other pathways that govern cell differentiation.

In SCLC, the pathway of differentiation has been relatively well defined. The neuroendocrine phenotype is dictated in part by the basic-helix-loop-helix (bHLH) transcription factor human achaete-scute homolog 1 (hASH1), which is expressed at abnormally high levels in SCLC associated with ectopic ACTH. The abnormal expression of hASH1 and other developmental transcription factors appears to provide a link between cell proliferation and differentiation.

Ectopic hormone production might be considered merely epiphomenon associated with cancer if it did not cause clinical manifestations. Excessive and unregulated production of hormones such as ACTH, PTHrP, and vasopressin can lead to substantial morbidity and complicate the cancer treatment plan. Moreover, the paraneoplastic endocrinopathies may be a presenting clinical feature of underlying malignancy and prompt the search for an unrecognized tumor.

A large number of paraneoplastic endocrine syndromes have been described, linking overproduction of particular hormones with specific types of tumors. However, certain recurring syndromes emerge from this group (**Table 93-1**). The most common paraneoplastic endocrine syndromes include hypercalcemia from overproduction of PTHrP and other factors, hyponatremia from excess vasopressin, and Cushing's syndrome from ectopic ACTH.

HYPERCALCEMIA CAUSED BY ECTOPIC PRODUCTION OF PTHRP

(See also [Chap. 410](#))

Etiology Humoral hypercalcemia of malignancy (HHM) occurs in up to 20% of patients with cancer. HHM is most common in cancers of the lung, head and neck, skin, esophagus, breast, and genitourinary tract and in multiple myeloma and lymphomas, as well as metastases associated with these, and other cancers. There are several distinct humoral causes of HHM, but it is caused most commonly by overproduction of PTHrP. In addition to acting as a circulating humoral factor, bone metastases (e.g., breast, multiple myeloma) may produce PTHrP and other chemokines, leading to local osteolysis and hypercalcemia. PTHrP may also affect the initiation and progression of tumors by acting through pro-survival and chemokine pathways.

PTHrP is structurally related to PTH and binds to the PTH receptor, explaining the similar biochemical features of HHM and hyperparathyroidism. PTHrP plays a key physiologic role in skeletal development and regulates cellular proliferation and differentiation in other tissues, including skin, bone marrow, breast, and hair follicles. The mechanism of PTHrP induction in malignancy is incompletely understood; however, tumor-bearing tissues commonly associated with HHM normally produce PTHrP during development or cell renewal. PTHrP expression is stimulated by hedgehog pathways and Gli transcription factors that are active in many malignancies. Transforming growth factor β (TGF- β), which is produced by many tumors, also stimulates PTHrP. Mutations in certain oncogenes, such as *Ras*, also can activate PTHrP expression, as does loss of the tumor suppressor, p53. In addition to its role in HHM, the PTHrP pathway may also provide a potential target for therapeutic intervention to impede cancer growth.

TABLE 93-1 Paraneoplastic Syndromes Caused by Ectopic Hormone Production

PARANEOPLASTIC SYNDROME	ECTOPIC HORMONE	TYPICAL TUMOR TYPES ^a
Common		
Hypercalcemia of malignancy	Parathyroid hormone-related protein (PTHrP) 1,25-Dihydroxyvitamin D Parathyroid hormone (PTH) (rare) Prostaglandin E ₂ (PGE ₂) (rare)	Squamous cell (head and neck, lung, skin), breast, genitourinary, gastrointestinal; osteolytic metastases Lymphomas Lung, ovary Renal, lung
Syndrome of inappropriate antidiuretic hormone secretion (SIADH)	Vasopressin	Lung (squamous, small cell), gastrointestinal, genitourinary, ovary
Cushing's syndrome	Adrenocorticotrophic hormone (ACTH) Corticotropin-releasing hormone (CRH) (rare) Ectopic expression of gastric inhibitory peptide (GIP), luteinizing hormone (LH)/human chorionic gonadotropin (hCG), other G protein-coupled receptors (rare)	Lung (small cell, bronchial carcinoid, adenocarcinoma, squamous), thymus, pancreatic islet, medullary thyroid carcinoma, pheochromocytoma Pancreatic islet, carcinoid, lung, prostate Macronodular adrenal hyperplasia
Less Common		
Non-islet cell hypoglycemia	Insulin-like growth factor type II (IGF-II) Insulin (rare)	Mesenchymal tumors, sarcomas, adrenal, hepatic, gastrointestinal, kidney, prostate Cervix (small-cell carcinoma)
Male feminization	hCG ^b	Testis (embryonal, seminomas), germinomas, choriocarcinoma, lung, hepatic, pancreatic islet
Diarrhea or intestinal hypermotility	Calcitonin ^c Vasoactive intestinal peptide (VIP)	Lung, colon, breast, medullary thyroid carcinoma Pancreas, pheochromocytoma, esophagus
Rare		
Oncogenic osteomalacia	Fibroblast growth factor 23 (FGF23) or phosphatonin	Hemangiopericytomas, osteoblastomas, fibromas, sarcomas, giant cell tumors, prostate, lung
Acromegaly	Growth hormone-releasing hormone (GHRH) Growth hormone (GH)	Pancreatic islet, bronchial, and other carcinoids Lung, pancreatic islet
Hyperthyroidism	Thyroid-stimulating hormone (TSH)	Hydatidiform mole, embryonal tumors, struma ovarii
Hypertension	Renin	Juxtaglomerular tumors, kidney, lung, pancreas, ovary
Consumptive hypothyroidism	Type 3 deiodinase	Hepatic and other hemangiomas

^aOnly the most common tumor types are listed. For most ectopic hormone syndromes, an extensive list of tumors has been reported to produce one or more hormones.

^bhCG is produced ectopically by trophoblastic tumors. Certain tumors produce disproportionate amounts of the hCG α or hCG β subunit. High levels of hCG rarely cause hyperthyroidism because of weak binding to the TSH receptor. ^cCalcitonin is produced ectopically by medullary thyroid carcinoma and is used as a tumor marker.

Another relatively common cause of HHM is excess production of 1,25-dihydroxyvitamin D. Like granulomatous disorders associated with hypercalcemia, lymphomas can produce an enzyme that converts 25-hydroxyvitamin D to the more active 1,25-dihydroxyvitamin D, leading to enhanced gastrointestinal calcium absorption. Other causes of HHM include tumor-mediated production of osteolytic cytokines and inflammatory mediators.

Clinical Manifestations The typical presentation of HHM is a patient with a known malignancy who is found to be hypercalcemic on routine laboratory tests. Less often, hypercalcemia is the initial presenting feature of malignancy. Particularly when calcium levels are markedly increased ($>3.5 \text{ mmol/L}$ [$>14 \text{ mg/dL}$]), patients may experience fatigue, mental status changes, polyuria, dehydration, or symptoms of nephrolithiasis. Hypercalcemia can shorten ST segments and QT intervals, as well as bundle branch blocks and bradycardia.

Diagnosis Features that favor HHM, as opposed to primary hyperparathyroidism, include known malignancy, recent onset of hypercalcemia, and very high serum calcium levels. Like hyperparathyroidism, hypercalcemia caused by PTHrP is accompanied by hypercalciuria and hypophosphatemia. Patients with HHM typically have metabolic alkalosis rather than hyperchloremic acidosis, as is seen in hyperparathyroidism. Measurement of PTH is useful to exclude primary hyperparathyroidism; the PTH level should be suppressed in HHM. An elevated PTHrP level confirms the diagnosis, and it is increased in ~80% of hypercalcemic patients with cancer. 1,25-Dihydroxyvitamin D levels may be increased in patients with lymphoma.

TREATMENT

Humoral Hypercalcemia of Malignancy

The management of HHM begins with removal of excess calcium in the diet, medications, or IV solutions. Saline rehydration (typically 200–500 mL/h) is used to dilute serum calcium and promote calciuresis; exercise caution in patients with cardiac, hepatic, or renal insufficiency. Forced diuresis with furosemide (20–80 mg IV in escalating doses) or other loop diuretics can enhance calcium excretion but provides relatively little value except in life-threatening hypercalcemia. When used, loop diuretics should be administered only after complete rehydration and with careful monitoring of fluid balance. Oral phosphorus (e.g., 250 mg Neutra-Phos 3–4 times daily) should be given until serum phosphorus is $>1 \text{ mmol/L}$ ($>3 \text{ mg/dL}$). Bisphosphonates such as pamidronate (60–90 mg IV), zoledronate (4–8 mg IV), and etidronate (7.5 mg/kg per day PO for 3–7 consecutive days) can reduce serum calcium within 1–2 days and suppress calcium release for several weeks. Bisphosphonate infusions can be repeated, or oral bisphosphonates can be used for chronic treatment. Denosumab (120 mg SC weekly for 4 weeks and then monthly) can be used in patients who do not respond adequately to bisphosphonates. It acts as a decoy receptor for RANK ligand to mitigate stimulation of osteoclasts. Cinacalcet (30 mg PO bid to 90 mg PO qid) stimulates calcium-sensing receptors to suppress PTH secretion and is therefore applicable in parathyroid carcinoma and rare cases of ectopic PTH-producing tumors. Hypercalcemia associated with lymphomas, multiple myeloma, or leukemia may respond to glucocorticoid treatment (e.g., prednisone 40–100 mg PO in four divided doses). Dialysis should be considered in severe hypercalcemia when saline hydration and bisphosphonate treatments are not possible or are too slow in onset. Previously used agents such as calcitonin and mithramycin have little utility now that bisphosphonates and other agents are available.

■ ECTOPIC VASOPRESSIN: TUMOR-ASSOCIATED SYNDROME OF INAPPROPRIATE ANTIDIURETIC HORMONE

(See also Chap. 53)

Etiology Vasopressin is an antidiuretic hormone normally produced by the posterior pituitary gland. Ectopic vasopressin production by tumors is a common cause of the syndrome of inappropriate antidiuretic hormone (SIADH), occurring in at least half of patients with SCLC. SIADH also can be caused by a number of nonneoplastic conditions, including central nervous system (CNS) trauma, infections, and medications (Chap. 381). Compensatory responses to SIADH, such as decreased thirst, may mitigate the development of hyponatremia. However, with prolonged production of excessive vasopressin, the osmostat controlling thirst and hypothalamic vasopressin secretion may become reset. In addition, intake of free water, orally or intravenously, can quickly worsen hyponatremia because of reduced renal diuresis.

Tumors with neuroendocrine features, such as SCLC and carcinoids, are the most common sources of ectopic vasopressin production, but it also occurs in other forms of lung cancer and with CNS lesions, head and neck cancer, and genitourinary, gastrointestinal, and ovarian cancers. The mechanism of activation of the vasopressin gene in these tumors is unknown, but the frequent concomitant expression of the adjacent oxytocin gene suggests derepression of this locus.

Clinical Manifestations Most patients with ectopic vasopressin secretion are asymptomatic and are identified because of the presence of hyponatremia on routine chemistry testing. Symptoms may include weakness, lethargy, nausea, confusion, depressed mental status, and seizures. The severity of symptoms reflects the rapidity of onset as well as the severity of hyponatremia. Hyponatremia usually develops slowly but may be exacerbated by the administration of IV fluids or the institution of new medications.

Diagnosis The diagnostic features of ectopic vasopressin production are the same as those of other causes of SIADH (Chaps. 53 and 381). Hyponatremia and reduced serum osmolality occur in the setting of an inappropriately normal or increased urine osmolality. Urine sodium excretion is normal or increased unless volume depletion is present. Other causes of hyponatremia should be excluded, including renal, adrenal, or thyroid insufficiency. Physiologic sources of vasopressin stimulation (CNS lesions, pulmonary disease, nausea), adaptive circulatory mechanisms (hypotension, heart failure, hepatic cirrhosis), and medications, including many chemotherapeutic agents, also should be considered as possible causes of hyponatremia. Vasopressin measurements are not usually necessary to make the diagnosis.

TREATMENT

Ectopic Vasopressin: Tumor-Associated SIADH

Most patients with ectopic vasopressin production develop hyponatremia over several weeks or months. The disorder should be corrected gradually unless mental status is altered or there is risk of seizures. Rapid correction can cause brain dehydration and central pontine myelinolysis. Treatment of the underlying malignancy may reduce ectopic vasopressin production, but this response is slow if it occurs at all. Fluid restriction to less than urine output, plus insensible losses, is often sufficient to correct hyponatremia partially. However, strict monitoring of the amount and types of liquids consumed or administered intravenously is required for fluid restriction to be effective. Salt tablets and saline are not helpful unless volume depletion is also present. Demeclocycline (150–300 mg orally 3–4 times daily) can be used to inhibit vasopressin action on the renal distal tubule, but its onset of action is relatively slow (1–2 weeks). The vaptan class of drugs acts by inhibiting vasopressin receptors (V_{1A} , V_{1B} , V_2) in the renal collecting ducts. Conivaptan, a nonpeptide V_2 -receptor antagonist, can be administered either PO (20–120 mg bid) or IV (10–40 mg) and is particularly effective when used in combination with fluid restriction in euvolemic hyponatremia. Tolvaptan (15 mg PO daily) is another vasopressin antagonist. The dose can be increased to 30–60 mg/d based on response. Severe hyponatremia ($\text{Na} < 115 \text{ meq/L}$) or mental status changes may require treatment with hypertonic (3%) or normal saline infusion together with furosemide to enhance free water clearance. The rate of sodium

correction should be slow (0.5–1 meq/L per hour) to prevent rapid fluid shifts and the possible development of central pontine myelinolysis.

CUSHING'S SYNDROME CAUSED BY ECTOPIC ACTH PRODUCTION

(See also [Chap. 386](#))

Etiology Ectopic ACTH production accounts for 10–20% of cases of Cushing's syndrome. The syndrome is particularly common in neuroendocrine tumors. SCLC is the most common cause of ectopic ACTH, followed by bronchial and thymic carcinoids, islet cell tumors, other carcinoids, and pheochromocytomas. Ectopic ACTH production is caused by increased expression of the proopiomelanocortin (*POMC*) gene, which encodes ACTH, along with melanocyte-stimulating hormone (MSH), β lipotropin, and several other peptides. In many tumors, there is abundant but aberrant expression of the *POMC* gene from an internal promoter, proximal to the third exon, which encodes ACTH. However, because this product lacks the signal sequence necessary for protein processing, it is not secreted. Increased production of ACTH arises instead from less abundant, but unregulated, *POMC* expression from the same promoter site used in the pituitary. Because tumors lack many of the enzymes needed to process the *POMC* polypeptide, it is typically released as multiple large, biologically inactive fragments along with relatively small amounts of fully processed, active ACTH.

Rarely, corticotropin-releasing hormone (CRH) is produced by pancreatic islet cell tumors, SCLC, medullary thyroid cancer, carcinoids, or prostate cancer. When levels are high enough, CRH can cause pituitary corticotrope hyperplasia and Cushing's syndrome. Tumors that produce CRH sometimes also produce ACTH, raising the possibility of a paracrine mechanism for ACTH production.

A distinct mechanism for ACTH-independent Cushing's syndrome involves ectopic expression of various G protein-coupled receptors in adrenal nodules. Ectopic expression of the gastric inhibitory peptide (GIP) receptor is the best-characterized example of this mechanism. In this case, meals induce GIP secretion, which inappropriately stimulates adrenal growth and glucocorticoid production.

Clinical Manifestations The clinical features of hypercortisolism are detected in only a fraction of patients with documented ectopic ACTH production. Patients with ectopic ACTH syndrome generally exhibit less marked weight gain and centripetal fat redistribution, probably because the exposure to excess glucocorticoids is relatively brief and because cachexia reduces the propensity for weight gain and fat deposition. The ectopic ACTH syndrome is associated with several clinical features that distinguish it from other causes of Cushing's syndrome (e.g., pituitary adenomas, adrenal adenomas, iatrogenic glucocorticoid excess). The metabolic manifestations of ectopic ACTH syndrome are dominated by fluid retention and hypertension, hypokalemia, metabolic alkalosis, glucose intolerance, and occasionally steroid psychosis. The very high ACTH levels often cause increased pigmentation, reflecting increased activity of MSH derived from the *POMC* precursor peptide. The extraordinarily high glucocorticoid levels in patients with ectopic sources of ACTH can lead to marked skin fragility and easy bruising. In addition, the high cortisol levels often overwhelm the renal 11 β -hydroxysteroid dehydrogenase type II enzyme, which normally inactivates cortisol and prevents it from binding to renal mineralocorticoid receptors. Consequently, in addition to the excess mineralocorticoids produced by ACTH stimulation of the adrenal gland, high levels of cortisol exert activity through the mineralocorticoid receptor, leading to severe hypokalemia.

Diagnosis The diagnosis of ectopic ACTH syndrome is usually not difficult in the setting of a known malignancy. Urine-free cortisol levels fluctuate but are typically greater than two to four times normal, and the plasma ACTH level is usually >22 pmol/L (>100 pg/mL). A suppressed ACTH level excludes this diagnosis and indicates an ACTH-independent cause of Cushing's syndrome (e.g., adrenal or exogenous glucocorticoid). In contrast to pituitary sources of ACTH,

most ectopic sources of ACTH do not respond to glucocorticoid suppression. Therefore, high-dose dexamethasone (8 mg PO) suppresses 8:00 A.M. serum cortisol (50% decrease from baseline) in ~80% of pituitary ACTH-producing adenomas but fails to suppress ectopic ACTH in ~90% of cases. Bronchial and other carcinoids are well-documented exceptions to these general guidelines, as these ectopic sources of ACTH may exhibit feedback regulation indistinguishable from pituitary adenomas, including suppression by high-dose dexamethasone, and ACTH responsiveness to adrenal blockade with metyrapone. If necessary, petrosal sinus catheterization can be used to evaluate a patient with ACTH-dependent Cushing's syndrome when the source of ACTH is unclear. After CRH stimulation, a 3:1 petrosal sinus:peripheral ACTH ratio strongly suggests a pituitary ACTH source. Imaging studies (computed tomography or magnetic resonance imaging) are also useful in the evaluation of suspected carcinoid lesions, allowing biopsy and characterization of hormone production using special stains. If available, positron emission tomography or octreotide scanning may identify some sources of ACTH production.

TREATMENT

Cushing's Syndrome Caused by Ectopic ACTH Production

The morbidity associated with the ectopic ACTH syndrome can be substantial. Patients may experience depression or personality changes because of extreme cortisol excess. Metabolic derangements, including diabetes mellitus and hypokalemia, can worsen fatigue. Poor wound healing and predisposition to infections can complicate the surgical management of tumors, and opportunistic infections caused by organisms such as *Pneumocystis carinii* and mycoses are often the cause of death in patients with ectopic ACTH production. These patients have increased risk of venous thromboembolism reflecting the combination of malignancy and altered coagulation factor profiles. Depending on prognosis and treatment plans for the underlying malignancy, measures to reduce cortisol levels are often indicated. Treatment of the underlying malignancy may reduce ACTH levels but is rarely sufficient to reduce cortisol levels to normal. Adrenalectomy is not practical for most of these patients but should be considered during surgery for the malignancy or if the underlying tumor is not resectable and the prognosis is otherwise favorable (e.g., carcinoid). Medical therapy with ketoconazole (300–600 mg PO bid), metyrapone (250–500 mg PO every 6 h), mitotane (3–6 g PO in four divided doses, tapered to maintain low cortisol production), etomidate (0.1–0.3 mg/kg/h IV), or other agents that block steroid synthesis or action is often the most practical strategy for managing the hypercortisolism associated with ectopic ACTH production. Glucocorticoid replacement should be provided to prevent adrenal insufficiency ([Chap. 386](#)). Unfortunately, many patients eventually progress despite medical blockade. Mifepristone (200–1000 mg PO qd) inhibits both glucocorticoid and progesterone receptors, has rapid onset of action, and improves glucose intolerance and hypertension in a subset of patients. ACTH-neutralizing antibodies and ACTH receptor blockers are under investigation, as are selective inhibitors of the glucocorticoid receptor.

TUMOR-INDUCED HYPOGLYCEMIA CAUSED BY EXCESS PRODUCTION OF INSULIN-LIKE GROWTH FACTOR TYPE II

(See also [Chap. 406](#)) Mesenchymal tumors, hemangiopericytomas, hepatocellular tumors, adrenal carcinomas, and a variety of other large tumors have been reported to produce excessive amounts of insulin-like growth factor type II (IGF-II) precursor, which binds weakly to insulin receptors and more strongly to IGF-I receptors, leading to insulin-like actions. The gene encoding IGF-II resides on chromosome 11p15, a locus that is normally imprinted (that is, expression

is exclusively from a single parental allele). Biallelic expression of the IGF-II gene occurs in a subset of tumors, suggesting loss of methylation and loss of imprinting as a mechanism for gene induction. In addition to increased IGF-II production, IGF-II bioavailability is increased due to complex alterations in circulating binding proteins. Increased IGF-II suppresses growth hormone (GH) and insulin, resulting in reduced IGF binding protein 3 (IGFBP-3), IGF-I, and acid-labile subunit (ALS). The reduction in ALS and IGFBP-3, which normally sequester IGF-II, causes it to be displaced to a small circulating complex that has greater access to insulin target tissues. For this reason, circulating IGF-II levels may not be markedly increased despite causing hypoglycemia. In addition to IGF-II-mediated hypoglycemia, tumors may occupy enough of the liver to impair gluconeogenesis.

In most cases, a tumor causing hypoglycemia is clinically apparent (usually >10 cm in size), and hypoglycemia develops in association with fasting. As with other causes of hypoglycemia, patients may present with sweating, tremors, palpitations, confusion, seizures, or coma. The diagnosis is made by documenting low serum glucose and suppressed insulin levels in association with symptoms of hypoglycemia. Serum IGF-II levels may not be increased (IGF-II assays may not detect IGF-II precursors), but an elevated IGF-II/IGF-I ratio greater than 10:1 is suggestive. Increased IGF-II mRNA expression is found in most of these tumors. Any medications associated with hypoglycemia should be eliminated. Treatment of the underlying malignancy, if possible, may reduce the predisposition to hypoglycemia. Frequent meals and IV glucose, especially during sleep or fasting, are often necessary to prevent hypoglycemia. Glucagon and glucocorticoids have also been used to enhance glucose production. Antibodies that inhibit IGF-II action are under development.

HUMAN CHORIONIC GONADOTROPIN

hCG is composed of α and β subunits and can be produced as intact hormone, which is biologically active, or as uncombined biologically inert subunits. Ectopic production of intact hCG occurs most often in association with testicular embryonal tumors, germ cell tumors, extragonadal germinomas, lung cancer, hepatoma, and pancreatic islet tumors. Ectopic production of hCG occurs with trophoblastic malignancies. hCG α subunit production is particularly common in lung cancer and pancreatic islet cancer. In men, high hCG levels stimulate steroidogenesis and aromatase activity in testicular Leydig cells, resulting in increased estrogen production and the development of gynecomastia. Precocious puberty in boys or gynecomastia in men should prompt measurement of hCG and consideration of a testicular tumor or another source of ectopic hCG production. Most women are asymptomatic. hCG is easily measured. Treatment should be directed at the underlying malignancy.

ONCOGENIC OSTEOMALACIA

Hypophosphatemic oncogenic osteomalacia, also called tumor-induced osteomalacia (TIO), is caused by excessive production of fibroblast growth factor 23 (FGF23), previously referred to as phosphatonin. Oncogenic osteomalacia is characterized by markedly reduced serum phosphorus and renal phosphate wasting, leading to muscle weakness, bone pain, and osteomalacia. Serum calcium and PTH levels are normal. FGF23 inhibits the renal conversion of 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D, resulting in low levels of 1,25-dihydroxyvitamin D. Oncogenic osteomalacia is usually caused by benign mesenchymal tumors, such as hemangiopericytomas, fibromas, and giant cell tumors, often of the skeletal extremities or head. It has also been described in sarcomas and in patients with prostate or lung cancer. Resection of the tumor reverses the disorder, confirming its humoral basis. FGF23 levels are increased in some, but not all, patients with osteogenic osteomalacia. FGF23 forms a ternary complex with the klotho protein and renal FGF receptors to reduce renal phosphate reabsorption. Treatment involves removal of the tumor, if possible, and supplementation with phosphate and vitamin D. Octreotide treatment reduces phosphate wasting in some patients with tumors that express somatostatin receptor subtype 2. Octreotide scans may also be useful in detecting these tumors. The calcium-sensing

receptor agonist, cinacalcet, has been effective in some patients, apparently by reducing PTH-mediated phosphaturia. FGF receptor inhibitors hold promise as future therapies targeted either to pathways that stimulate FGR23 production (e.g., FGFR1) or inhibit its action (e.g., FGF23 receptor).

CONSUMPTIVE HYPOTHYROIDISM

Newborns with hepatic hemangiomas can develop a rare form of hypothyroidism caused by overexpression of type 3 deiodinase (D3), an enzyme that degrades and inactivates thyroxine (T_4) and triiodothyronine (T_3). The very high expression of D3 and consumption of thyroid hormones apparently outstrip the thyroid gland's rate of hormone production. The disorder is characterized by low T_4 , low T_3 , high TSH, and markedly elevated reverse T_3 (rT_3), reflecting the degradation of T_4 to rT_3 . In addition to treating the underlying hemangioma (rarely other tumor types), patients are treated with L-thyroxine replacement, titrated to normalize TSH. Steroids and propranolol may provide benefit, perhaps by inhibiting growth factor pathways thought to stimulate D3 production.

HEMATOLOGIC SYNDROMES

The elevation of granulocyte, platelet, and eosinophil counts in most patients with myeloproliferative disorders is caused by the proliferation of the myeloid elements due to the underlying disease rather than to a paraneoplastic syndrome. The paraneoplastic hematologic syndromes in patients with solid tumors are less well characterized than are the endocrine syndromes because the ectopic hormone(s) or cytokines responsible have not been identified in most of these tumors (Table 93-2). The extent of the paraneoplastic syndromes parallels the course of the cancer. With very rare exception, red cell, white cell or platelet numbers are self-limited and not associated with symptomatic abnormalities. In some circumstances, elevations in platelet counts can be a marker that influences prognosis. By far, the most consequential hematologic abnormality in cancer patients is hypercoagulability.

ERYTHROCYTOSIS

Ectopic production of erythropoietin by cancer cells causes most paraneoplastic erythrocytosis. The ectopically produced erythropoietin stimulates the production of red blood cells (RBCs) in the bone marrow and raises the hematocrit. Other lymphokines and hormones produced by cancer cells may stimulate erythropoietin release but have not been proved to cause erythrocytosis.

Most patients with erythrocytosis have an elevated hematocrit (>52% in men, >48% in women) that is detected on a routine blood

TABLE 93-2 Paraneoplastic Hematologic Syndromes

SYNDROME	PROTEINS	CANCERS TYPICALLY ASSOCIATED WITH SYNDROME
Erythrocytosis	Erythropoietin	Renal cancers, hepatocarcinoma, cerebellar hemangioblastomas
Granulocytosis	G-CSF, GM-CSF, IL-6	Lung cancer, gastrointestinal cancer, ovarian cancer, genitourinary cancer, Hodgkin's disease
Thrombocytosis	IL-6	Lung cancer, gastrointestinal cancer, breast cancer, ovarian cancer, lymphoma
Eosinophilia	IL-5	Lymphoma, leukemia, lung cancer
Thrombophlebitis	Unknown	Lung cancer, pancreatic cancer, gastrointestinal cancer, breast cancer, genitourinary cancer, ovarian cancer, prostate cancer, lymphoma

Abbreviations: G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; IL, interleukin.

count. Approximately 3% of patients with renal cell cancer, 10% of patients with hepatoma, and 15% of patients with cerebellar hemangioblastomas have erythrocytosis. In most cases, the erythrocytosis is asymptomatic.

Patients with erythrocytosis due to a renal cell cancer, hepatoma, or CNS cancer should have measurement of red cell mass. If the red cell mass is elevated, the serum erythropoietin level should be measured. Patients with a cancer that has been associated with erythrocytosis, elevated erythropoietin levels, and no other explanation for erythrocytosis (e.g., hemoglobinopathy that causes increased O₂ affinity; Chaps. 63 and 98) have the paraneoplastic syndrome.

TREATMENT

Erythrocytosis

Successful resection of the cancer usually resolves the erythrocytosis. If the tumor cannot be resected or treated effectively with radiation therapy or chemotherapy, phlebotomy may control any symptoms or risk related to erythrocytosis.

■ GRANULOCYTOSIS

Approximately 30% of patients with solid tumors have granulocytosis (granulocyte count >8000/ μ L). In about half of patients with granulocytosis and cancer, the granulocytosis has an identifiable nonparaneoplastic etiology (e.g., infection, tumor necrosis, glucocorticoid administration). The other patients have proteins in urine and serum that stimulate the growth of bone marrow cells. Tumors and tumor cell lines from patients with lung, ovarian, and bladder cancers have been documented to produce granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), and/or interleukin 6 (IL-6). However, the etiology of granulocytosis has not been characterized in most patients.

Patients with granulocytosis are nearly all asymptomatic, and the differential white blood cell count does not have a shift to immature forms of neutrophils. Granulocytosis occurs in 40% of patients with lung and gastrointestinal cancers, 20% of patients with breast cancer, 30% of patients with brain tumors and ovarian cancers, 20% of patients with Hodgkin's disease, and 10% of patients with renal cell carcinoma. Patients with advanced-stage disease are more likely to have granulocytosis than are those with early-stage disease.

Paraneoplastic granulocytosis does not require treatment. The granulocytosis resolves when the underlying cancer is treated.

■ THROMBOCYTOSIS

Some 35% of patients with thrombocytosis (platelet count >400,000/ μ L) have an underlying diagnosis of cancer. IL-6, a candidate molecule for the etiology of paraneoplastic thrombocytosis, stimulates the production of platelets *in vitro* and *in vivo*. Some patients with cancer and thrombocytosis have elevated levels of IL-6 in plasma. Another candidate molecule is thrombopoietin, a peptide hormone that stimulates megakaryocyte proliferation and platelet production. The etiology of thrombocytosis has not been established in most cases.

Patients with thrombocytosis are nearly all asymptomatic. Thrombocytosis is not clearly linked to thrombosis in patients with cancer. Thrombocytosis is present in 40% of patients with lung and gastrointestinal cancers; 20% of patients with breast, endometrial, and ovarian cancers; and 10% of patients with lymphoma. Patients with thrombocytosis are more likely to have advanced-stage disease and have a poorer prognosis than do patients without thrombocytosis. In ovarian cancer, IL-6 has been shown to directly promote tumor growth. Paraneoplastic thrombocytosis does not require treatment other than treatment of the underlying tumor.

■ EOSINOPHILIA

Eosinophilia is present in ~1% of patients with cancer. Tumors and tumor cell lines from patients with lymphomas or leukemia may

produce IL-5, which stimulates eosinophil growth. Activation of IL-5 transcription in lymphomas and leukemias may involve translocation of the long arm of chromosome 5, to which the genes for IL-5 and other cytokines map.

Patients with eosinophilia are typically asymptomatic. Eosinophilia is present in 10% of patients with lymphoma, 3% of patients with lung cancer, and occasional patients with cervical, gastrointestinal, renal, and breast cancer. Patients with markedly elevated eosinophil counts (>5000/ μ L) can develop shortness of breath and wheezing. A chest radiograph may reveal diffuse pulmonary infiltrates from eosinophil infiltration and activation in the lungs.

TREATMENT

Eosinophilia

Definitive treatment is directed at the underlying malignancy. Tumors should be resected or treated with radiation or chemotherapy. In most patients who develop shortness of breath related to eosinophilia, symptoms resolve with the use of oral or inhaled glucocorticoids. IL-5 antagonists exist but have not been evaluated in this clinical setting.

■ THROMBOPHLEBITIS AND DEEP VENOUS THROMBOSIS

Deep venous thrombosis and pulmonary embolism are the most common thrombotic conditions in patients with cancer. Migratory or recurrent thrombophlebitis may be the initial manifestation of cancer. Nearly 15% of patients who develop deep venous thrombosis or pulmonary embolism have a diagnosis of cancer (Chap. 117). The coexistence of peripheral venous thrombosis with visceral carcinoma, particularly pancreatic cancer, is called *Trousseau's syndrome*.

Pathogenesis Patients with cancer are predisposed to thromboembolism because they are often at bed rest or immobilized, and tumors may obstruct or slow blood flow. Postoperative deep venous thrombosis is twice as common in cancer patients who undergo surgery. Chronic IV catheters also predispose to clotting. In addition, clotting may be promoted by release of procoagulants or cytokines from tumor cells or associated inflammatory cells or by platelet adhesion or aggregation. The specific molecules that promote thromboembolism have not been identified.

Chemotherapeutic agents, particularly those associated with endothelial damage, can induce venous thrombosis. The annual risk of venous thrombosis in patients with cancer receiving chemotherapy is about 11%, sixfold higher than the risk in the general population. Bleomycin, L-asparaginase, nitrogen mustard, thalidomide analogues, cisplatin-based regimens, and high doses of busulfan and carmustine are all associated with an increased risk.

In addition to cancer and its treatment causing secondary thrombosis, primary thrombophilic diseases may be associated with cancer. For example, the antiphospholipid antibody syndrome is associated with a wide range of pathologic manifestations (Chap. 357). About 20% of patients with this syndrome have cancers. Among patients with cancer and antiphospholipid antibodies, 35–45% develop thrombosis.

Clinical Manifestations Patients with cancer who develop deep venous thrombosis usually develop swelling or pain in the leg, and physical examination reveals tenderness, warmth, and redness. Patients who present with pulmonary embolism develop dyspnea, chest pain, and syncope, and physical examination shows tachycardia, cyanosis, and hypotension. Some 5% of patients with no history of cancer who have a diagnosis of deep venous thrombosis or pulmonary embolism will have a diagnosis of cancer within 1 year. The most common cancers associated with thromboembolic episodes include lung, pancreatic, gastrointestinal, breast, ovarian, and genitourinary cancers; lymphomas; and brain tumors. Patients with cancer who undergo

surgical procedures requiring general anesthesia have a 20–30% risk of deep venous thrombosis.

Diagnosis The diagnosis of deep venous thrombosis in patients with cancer is made by impedance plethysmography or bilateral compression ultrasonography of the leg veins. Patients with a noncompressible venous segment have deep venous thrombosis. If compression ultrasonography is normal and there is a high clinical suspicion for deep venous thrombosis, venography should be done to look for a luminal filling defect. Elevation of D-dimer is not as predictive of deep venous thrombosis in patients with cancer as it is in patients without cancer; elevations are seen in people over age 65 years without concomitant evidence of thrombosis, probably as a consequence of increased thrombin deposition and turnover in aging.

Patients with symptoms and signs suggesting a pulmonary embolism should be evaluated with a chest radiograph, electrocardiogram, arterial blood gas analysis, and ventilation-perfusion scan. Patients with mismatched segmental perfusion defects have a pulmonary embolus. Patients with equivocal ventilation-perfusion findings should be evaluated as described above for deep venous thrombosis in their legs. If deep venous thrombosis is detected, they should be anticoagulated. If deep venous thrombosis is not detected, they should be considered for a pulmonary angiogram.

Patients without a diagnosis of cancer who present with an initial episode of thrombophlebitis or pulmonary embolus need no additional tests for cancer other than a careful history and physical examination. In light of the many possible primary sites, diagnostic testing in asymptomatic patients is wasteful. However, if the clot is refractory to standard treatment or is in an unusual site, or if the thrombophlebitis is migratory or recurrent, efforts to find an underlying cancer are indicated.

TREATMENT

Thrombophlebitis and Deep Venous Thrombosis

Patients with cancer and a diagnosis of deep venous thrombosis or pulmonary embolism should be treated initially with IV unfractionated heparin or low-molecular-weight heparin for at least 5 days, and warfarin should be started within 1 or 2 days. The warfarin dose should be adjusted so that the international normalized ratio (INR) is 2–3. Patients with proximal deep venous thrombosis and a relative contraindication to heparin anticoagulation (hemorrhagic brain metastases or pericardial effusion) should be considered for placement of a filter in the inferior vena cava (Greenfield filter) to prevent pulmonary embolism. Warfarin should be administered for 3–6 months. An alternative approach is to use low-molecular-weight heparin for 6 months. The new oral anticoagulants (factor Xa and thrombin inhibitors) are attractive because they do not require close monitoring of the prothrombin time and are not affected by dietary factors. Oral apixaban (10 mg bid for 7 days followed by 5 mg bid for 6 months) is noninferior to dalteparin in the treatment of cancer patients who develop deep vein thrombosis or pulmonary embolism. Patients with cancer who undergo a major surgical procedure should be considered for heparin prophylaxis or pneumatic boots. Breast cancer patients undergoing chemotherapy and patients with implanted catheters should be considered for prophylaxis. Guidelines recommend that hospitalized patients with cancer and patients receiving a thalidomide analogue receive prophylaxis with low-molecular-weight heparin or low-dose aspirin. Use of prophylaxis routinely during chemotherapy is controversial. Risk is affected by type of cancer, type of therapy, blood counts, and body mass index (all taken into account in the Khorana risk score; **Table 93-3**). Studies of Khorana high-risk patients with cancer using rivaroxaban and apixaban as clot prophylaxis have resulted in a 50% reduction in risk with a level of bleeding of about 5%. However,

TABLE 93-3 Khorana Risk Score for Venous Thromboembolism in Cancer Patients

PATIENT CHARACTERISTICS	RISK SCORE POINTS	
Site of cancer		
Very high risk (stomach, pancreas)	2	
High risk (lung, lymphoma, gynecologic, genitourinary excluding prostate)	1	
Prechemotherapy platelet count $\geq 350,000/\mu\text{L}$	1	
Hemoglobin level $< 10 \text{ g/dL}$ or use of red cell growth factors	1	
Prechemotherapy leukocyte count $> 11,000/\mu\text{L}$	1	
BMI $\geq 35 \text{ kg/m}^2$	1	
RISK SCORE (POINTS)	RISK CATEGORY	RATES OF sVTE ACCORDING TO SCORES (%)
0	Low	0.3–0.8
1–2	Intermediate	1.8–2.0
≥ 3	High	6.7–7.1

Abbreviations: BMI, body mass index; sVTE, symptomatic venous thromboembolism.

Source: AJ Muñoz Martín et al: Clinical guide SEOM on venous thromboembolism in cancer patients. Clin Transl Oncol 16:1079, 2014.

prophylaxis is not routinely recommended by the American Society of Clinical Oncology.

MISCELLANEOUS REMOTE EFFECTS OF CANCER

Patients with cancer can develop paraneoplastic autoimmune disorders (e.g., thrombocytopenia) and dysfunction of organs not directly invaded or involved with the cancer (rheumatologic and renal abnormalities are among the most frequent). The pathogenesis of these disorders is undefined, but often, the conditions reverse if the tumor is removed or successfully treated.

Cutaneous paraneoplastic syndromes are discussed in [Chap. 58](#). Neurologic paraneoplastic syndromes are discussed in [Chap. 94](#).

FURTHER READING

- AGNELLI G et al: Apixaban for the treatment of venous thromboembolism associated with cancer. N Engl J Med 382:1599, 2020.
- ASONITIS N et al: Diagnosis, pathophysiology and management of hypercalcemia in malignancy: A review of the literature. Horm Metab Res 51:770, 2019.
- CATANI MV et al: The “Janus face” of platelets in cancer. Int J Mol Sci 21:788, 2020.
- DYNKEVICH Y et al: Tumors, IGF-2, and hypoglycemia: insights from the clinic, the laboratory, and the historical archive. Endocr Rev 34:798, 2013.
- ELLISON DH, BERL T: The syndrome of inappropriate antidiuresis. N Engl J Med 356:2064, 2007.
- FEELDERS RA et al: Advances in the medical treatment of Cushing’s syndrome. Lancet Diabetes Endocrinol 7:300, 2019.
- HARTLEY IR et al: Targeted FGFR blockade for the treatment of tumor-induced osteomalacia. N Engl J Med 383:14, 2020.
- ISIDORI AM et al: The ectopic adrenocorticotropin syndrome: Clinical features, diagnosis, management and long-term follow-up. J Clin Endocrinol Metab 91:371, 2006.
- LIN RJ et al: Paraneoplastic thrombocytosis: The secrets of tumor self-promotion. Blood 124:184, 2014.
- PELOSOF LC, GERBER DE: Paraneoplastic syndromes: An approach to diagnosis and treatment. Mayo Clin Proc 85:838, 2010.
- WORKENEH BT et al: Hyponatremia in the cancer patient. Kidney Int 98:870, 2020.

Paraneoplastic Neurologic Syndromes and Autoimmune Encephalitis

Josep Dalmau, Myrna R. Rosenfeld,
Francesc Graus

Paraneoplastic neurologic disorders (PNDs) are cancer-related syndromes that can affect any part of the nervous system (**Table 94-1**). They are caused by mechanisms other than metastasis or by any of the complications of cancer such as coagulopathy, stroke, metabolic and nutritional conditions, infections, and side effects of cancer therapy. In 60% of patients, the neurologic symptoms precede the cancer diagnosis. Clinically disabling PNDs occur in 0.5–1% of all cancer patients, but they affect 2–3% of patients with neuroblastoma or small-cell lung cancer (SCLC) and 30–50% of patients with thymoma.

PATHOGENESIS

Most PNDs are mediated by immune responses triggered by neuronal proteins ectopically expressed by tumors (e.g., SCLC and other cancers) or as a result of altered immunologic responses caused by some types of tumors such as thymomas or lymphomas. In PNDs of the central nervous system (CNS), many antibody-associated immune responses have been identified (**Table 94-2**). These antibodies react with neurons and the patient's tumor, and their detection in serum or cerebrospinal fluid (CSF) usually predicts the presence of cancer. When the antigens are intracellular, most syndromes are associated with extensive infiltrates of CD4+ and CD8+ T cells, microglial activation, gliosis, and variable neuronal loss. The infiltrating T cells are often in close contact with neurons undergoing degeneration, suggesting a primary pathogenic role. T-cell-mediated cytotoxicity may contribute directly to cell death in these PNDs and probably underlies the resistance of many of these conditions to therapy.

In contrast to the predominant role of cytotoxic T-cell mechanisms in PND associated with antibodies against intracellular antigens, those associated with antibodies to antigens expressed on the neuronal cell surface of the CNS or at the neuromuscular junction are mediated by

TABLE 94-2 Antibodies to Intracellular Antigens, Syndromes, and Associated Cancers

ANTIBODY	ASSOCIATED NEUROLOGIC SYNDROME(S)	TUMORS
Anti-Hu (ANNA1)	Encephalomyelitis, subacute sensory neuronopathy	SCLC
Anti-Yo (PCA1)	Cerebellar degeneration	Ovary, breast
Anti-Ri (ANNA2)	Cerebellar degeneration, opsoclonus, brainstem encephalitis	Breast, gynecologic, SCLC
Anti-CRMP5 (CV2)	Encephalomyelitis, chorea, optic neuritis, uveitis, peripheral neuropathy	SCLC, thymoma, other
Anti-Ma proteins	Limbic, hypothalamic, brainstem encephalitis	Testicular (Ma2), other (Ma)
Anti-Kelch-like protein 11	Brainstem encephalitis, ataxia, hearing loss, diplopia	Seminoma, germ-cell tumor, teratoma
Anti-amphiphysin ^a	Stiff-person syndrome, encephalomyelitis	Breast, SCLC
Recoverin, bipolar cell antibodies, others ^b	Cancer-associated retinopathy (CAR), melanoma-associated retinopathy (MAR)	SCLC (CAR), melanoma (MAR)
Anti-GAD	Stiff-person, cerebellar syndromes, limbic encephalitis	Infrequent tumor association (thymoma and several cancers)

^aAmphiphysin is likely exposed to the cell surface during synaptic vesicle endocytosis. ^bA variety of target antigens have been identified.

Abbreviations: CRMP, collapsin response-mediator protein; SCLC, small-cell lung cancer.

direct antibody effects on the target antigens and are more responsive to immunotherapy (**Table 94-3**, **Fig. 94-1**). These disorders occur with and without a cancer association and may affect children and young adults. Some disorders are triggered by viral encephalitis such as herpes simplex virus encephalitis or Japanese encephalitis, leading to autoimmune encephalitis.

In patients with cancer, the use of immune checkpoint inhibitors is associated in rare instances with immune-related adverse events accompanied by neuronal antibodies, which are indistinguishable from paraneoplastic neurologic syndromes.

Other PNDs are likely immune-mediated, although their antigens are unknown. The best example is opsoclonus-myoclonus syndrome associated with neuroblastoma or SCLC. For still other PNDs, the cause remains quite obscure. These include, among others, several neuropathies that occur in the terminal stages of cancer and a number of neuropathies associated with plasma cell dyscrasias or lymphoma without evidence of tumor infiltration or deposits of immunoglobulin, cryoglobulin, or amyloid.

TABLE 94-1 Paraneoplastic Syndromes of the Nervous System

CLASSIC SYNDROMES: USUALLY OCCUR WITH CANCER ASSOCIATION	NONCLASSIC SYNDROMES: MAY OCCUR WITH AND WITHOUT CANCER ASSOCIATION
Encephalomyelitis	Brainstem encephalitis
Limbic encephalitis	Stiff-person syndrome
Cerebellar degeneration (adults)	Progressive encephalomyelitis with rigidity and myoclonus
Opsoclonus-myoclonus	Necrotizing myelopathy
Subacute sensory neuronopathy	Motor neuron disease
Gastrointestinal paresis or pseudo-obstruction	Guillain-Barré syndrome
Dermatomyositis (adults)	Subacute and chronic mixed sensory-motor neuropathies
Lambert-Eaton myasthenic syndrome	Neuropathy associated with plasma cell dyscrasias and lymphoma
Cancer- or melanoma-associated retinopathy	Vasculitis of nerve
	Pure autonomic neuropathy
	Acute necrotizing myopathy
	Polymyositis
	Optic neuropathy
	BDUMP
	Peripheral nerve hyperexcitability (neuromyotonia)
	Myasthenia gravis

Abbreviation: BDUMP, bilateral diffuse uveal melanocytic proliferation.

APPROACH TO THE PATIENT

Paraneoplastic Neurologic Disorders

Three key concepts are important for the diagnosis and management of PNDs. First, it is common for symptoms to appear before the presence of a tumor is known; second, the neurologic syndrome usually develops rapidly, producing severe deficits in a short period of time; and third, there is evidence that prompt tumor control improves the neurologic outcome. Therefore, the major concern of the physician is to recognize a disorder promptly as paraneoplastic and to identify and treat the tumor.

PND OF THE CENTRAL NERVOUS SYSTEM AND DORSAL ROOT GANGLIA

When symptoms involve brain, spinal cord, or dorsal root ganglia, the suspicion of PND is usually based on a combination of clinical, radiologic, and CSF findings. Presence of antineuronal antibodies

TABLE 94-3 Antibodies to Cell Surface or Synaptic Antigens, Syndromes, and Associated Tumors

ANTIBODY	NEUROLOGIC SYNDROME	TUMOR TYPE WHEN ASSOCIATED
Anti-NMDAR ^a	Anti-NMDAR encephalitis	Teratoma in young women (children and men rarely have tumors)
Anti-AMPAR ^a	Limbic encephalitis with relapses	SCLC, thymoma, breast, in ~70% of the patients
Anti-GluK2 ^a	Encephalitis, cerebellar ataxia, cerebellitis	No tumor, rarely teratoma
Anti-GABA _A R ^a	Encephalitis with prominent seizures and status epilepticus	Thymoma in ~30% of the patients
Anti-GABA _B R ^b	Limbic encephalitis with early and prominent seizures	SCLC in ~50% of the patients
Glycine receptor ^a	PERM, stiff-person syndrome	Rarely, thymoma, lung, Hodgkin's
Anti-mGluR5 ^a	Autoimmune encephalitis without distinctive features	Hodgkin's lymphoma, or no tumor
Anti-dopamine-2R ^b	Basal ganglia encephalitis	No cancer association
Anti-LGI1 ^{a,c}	Limbic encephalitis, hyponatremia, facioabdominal dystonic seizures	Rarely thymoma
Anti-Caspr2 ^{a,c}	Limbic encephalitis, ataxia, peripheral nerve hyperexcitability, neuropathy, Morvan's syndrome	~20% thymoma. In cases of Morvan syndrome: ~40% thymoma
Anti-DPPX ^a	Agitation, myoclonus, tremor, seizures, hyperekplexia, encephalomyelitis with rigidity	No cancer, but frequent diarrhea or cachexia suggesting paraneoplasia
Anti-neurexin 3 α ^b	Autoimmune encephalitis without distinctive features	No cancer association
IgLON5 ^a	NREM and REM sleep disorder, brainstem dysfunction, movement disorder, obstructive sleep apnea, stridor	No tumor association
Anti-mGluR1 ^a	Cerebellar syndrome	Hodgkin's lymphoma, or no tumor
Anti-mGluR2	Cerebellar syndrome	Small-cell neuroendocrine tumor, rhabdomyosarcoma
Anti-Tr (DNER)	Cerebellar syndrome	Hodgkin's lymphoma, or no tumor
Anti-Sez6l2	Cerebellar ataxia, postural instability, frequent falls, dysarthria, extrapyramidal symptoms	No cancer association
Anti-MOG	ADEM, optic neuritis, myelitis, cortical encephalitis	No cancer association
Anti-AChR (muscle) ^a	Myasthenia gravis	Thymoma
Anti-AChR (neuronal) ^a	Autonomic ganglionopathy	SCLC
Anti-VGCC ^a	LEMS, cerebellar degeneration	SCLC

^aA direct pathogenic role of these antibodies has been demonstrated in cultured neurons or animal models. ^bThese antibodies are strongly suspected to be pathogenic. ^cPreviously named voltage-gated potassium channel antibodies (VGKC); currently included under the term VGKC-complex proteins. Of note, the significance of antibodies to VGKC-complex proteins other than LGI1 and Caspr2 is uncertain (the antigens are unknown, and the response to immunotherapy is variable).

Abbreviations: AChR, acetylcholine receptor; ADEM, acute disseminated encephalomyelitis; AMPAR, α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor; Caspr2, contactin-associated protein-like 2; DNER, delta/notch-like epidermal growth factor-related receptor; DPPX, dipeptidyl-peptidase-like protein-6; GABA_AR, γ -aminobutyric acid B receptor; GAD, glutamic acid decarboxylase; GluK2, glutamate receptor ionotropic kainate 2; mGluR, metabotropic glutamate receptor; LEMS, Lambert-Eaton myasthenic syndrome; LGI1, leucine-rich glioma-inactivated 1; MOG, myelin oligodendrocyte glycoprotein; NMDAR, *N*-methyl-D-aspartate receptor; NREM, non-rapid eye movement; PERM, progressive encephalomyelitis with rigidity and myoclonus; REM, rapid eye movement; SCLC, small-cell lung cancer; Sez6l2, Seizure-related 6 homolog like 2; VGCC, voltage-gated calcium channel.

(Tables 94-2 and 94-3) may help in the diagnosis, but only 60–70% of PNDs of the CNS and <20% of those involving the peripheral nervous system have neuronal or neuromuscular junction antibodies that can be used as diagnostic tests.

Magnetic resonance imaging (MRI) and CSF studies are important to rule out neurologic complications due to the direct spread of cancer, particularly metastatic and leptomeningeal disease. In most PNDs, the MRI findings are nonspecific. Paraneoplastic limbic encephalitis is usually associated with characteristic MRI abnormalities in the mesial temporal lobes (see below), but similar findings can occur with other disorders (e.g., nonparaneoplastic autoimmune limbic encephalitis and human herpesvirus type 6 [HHV-6] encephalitis) (Fig. 94-2A). The CSF profile of patients with PND of the CNS or dorsal root ganglia typically consists of mild to moderate pleocytosis (<200 mononuclear cells, predominantly lymphocytes), an increase in the protein concentration, and a variable presence of oligoclonal bands. There are no specific electrophysiologic tests that are diagnostic of PND. Moreover, a biopsy of the affected tissue is often difficult to obtain, and although useful to rule out other disorders (e.g., metastasis), the pathologic findings are not specific for PND.

PND OF NERVE AND MUSCLE

If symptoms involve peripheral nerve, neuromuscular junction, or muscle, the diagnosis of a specific PND is usually established on clinical, electrophysiologic, and pathologic grounds. The clinical history, accompanying symptoms (e.g., anorexia, weight loss), and type of syndrome dictate the studies and degree of effort needed to demonstrate a neoplasm. For example, the frequent association of Lambert-Eaton myasthenic syndrome (LEMS) with SCLC should lead to a chest and abdomen computed tomography (CT) or body positron emission tomography (PET) scan and, if negative, periodic tumor screening for at least 3 years after the neurologic diagnosis. In contrast, the weak association of polymyositis with cancer calls into question the need for repeated cancer screenings in this situation. Serum and urine immunofixation studies should be considered in patients with peripheral neuropathy of unknown cause; detection of a monoclonal gammopathy suggests the need for additional studies to uncover a B-cell or plasma-cell malignancy. In paraneoplastic neuropathies, diagnostically useful antineuronal antibodies are limited to CRMP5 (CV2) and Hu (ANNA1).

For any type of PND, if antineuronal antibodies are negative, the diagnosis relies on the demonstration of cancer and the exclusion of other cancer-related or independent neurologic disorders. Combined CT and PET scans often uncover tumors undetected by other tests. For germ cell tumors of the testis and teratomas of the ovary, ultrasound (testicular, transvaginal, or pelvic) and MRI or CT of the abdomen and pelvis may reveal tumors undetectable by PET.

SPECIFIC PARANEOPLASTIC NEUROLOGIC SYNDROMES

■ PARANEOPLASTIC ENCEPHALOMYELITIS AND FOCAL ENCEPHALITIS WITH ANTIBODIES AGAINST INTRACELLULAR NEURONAL PROTEINS

The term *encephalomyelitis* describes an inflammatory process with multifocal involvement of the nervous system, including brain, brainstem, cerebellum, and spinal cord. It is often associated with dorsal root ganglia and autonomic dysfunction. For any given patient, the clinical manifestations are determined by the areas predominantly involved, but pathologic studies almost always reveal abnormalities beyond the symptomatic regions. Several clinicopathologic syndromes may occur alone or in combination: (1) *cortical encephalitis*, which may present as “epilepsia partialis continua”; (2) *limbic encephalitis*, characterized by confusion, depression, agitation, anxiety, severe deficit forming new memories (“short-term memory deficit”), and temporal lobe or generalized seizures (the MRI usually shows unilateral or bilateral medial temporal lobe abnormalities, best seen with T2 and fluid-attenuated

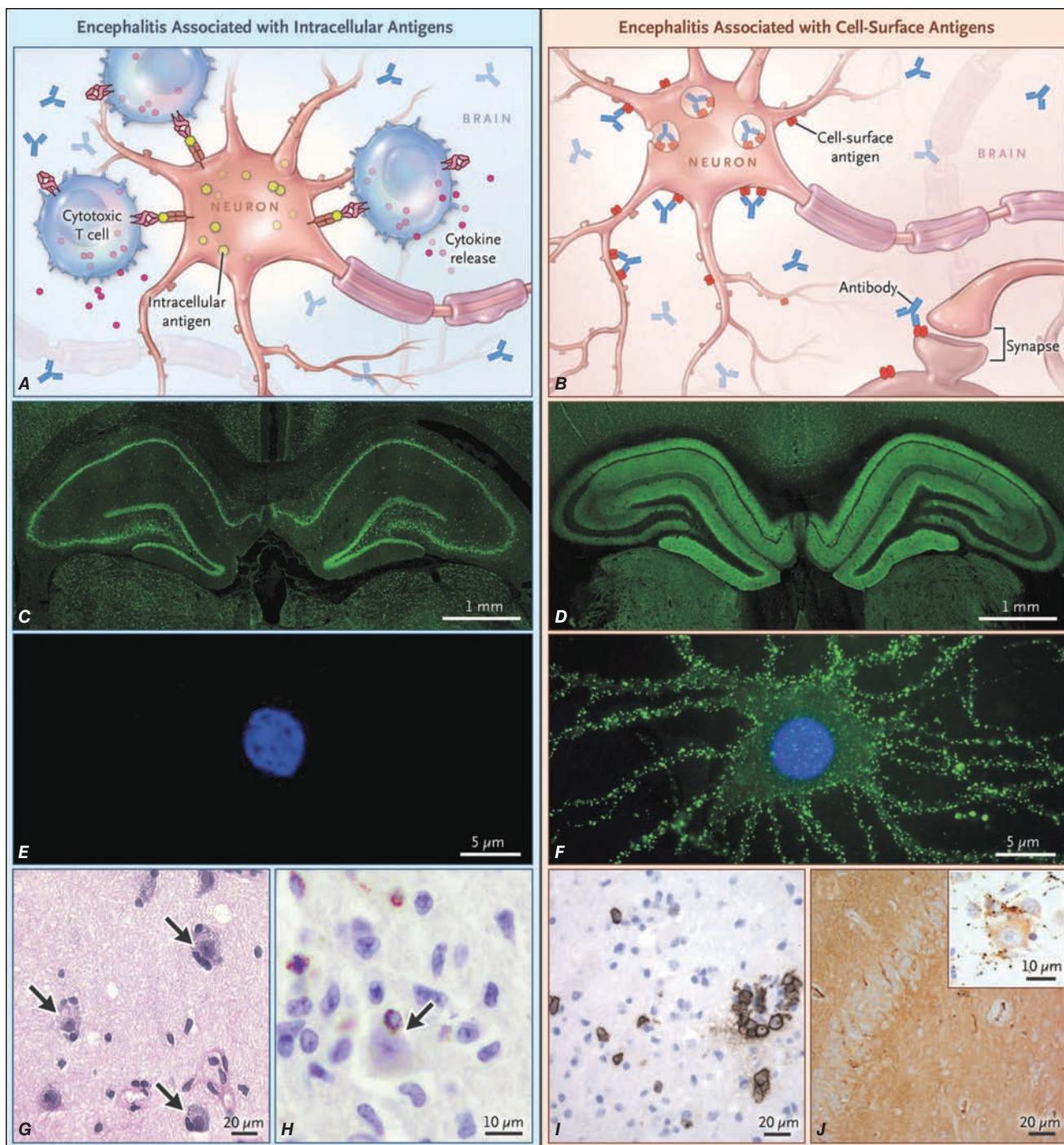


FIGURE 94-1 Antibody reactivity and pathologic findings in patients with antibodies against intracellular antigens compared with those of patients with antibodies against neuronal surface antigens. In encephalitis associated with antibodies against intracellular antigens, the antibodies cannot reach the intracellular epitopes and cytotoxic T-cell mechanisms are predominantly involved (**A**), whereas in encephalitis with antibodies against surface antigens, the antibodies have access to the epitopes and can potentially alter the structure and function of the antigen (**B**). The Hu antibodies (**C, E**) are shown here to exemplify the group of antibodies against intracellular antigens, and the NMDAR antibodies (**D, F**) are shown to exemplify the group of antibodies against cell-surface antigens. In rodent brain immunofluorescence with tissue permeabilized to allow entry of antibodies, the Hu antibodies produce a discrete pattern of cellular immunolabeling (**C**), whereas the NMDAR antibodies produce a pattern of neuropil-like immunolabeling (**D**). In contrast, with live cultured neurons, only the NMDAR antibodies have access to the target antigen showing intense immunolabeling (**F**), whereas the Hu antibodies cannot reach the intracellular antigen showing no immunolabeling (**E**). In autopsy studies, patients with encephalitis associated with antibodies to intracellular antigens (Hu or other) have extensive neuronal loss and inflammatory infiltrates (not shown); the T cells show direct contact with neurons (arrows in **G**) likely contributing to neuronal degeneration via perforin and granzyme mechanisms (arrow in **H**). In contrast, patients with antibodies against cell-surface antigens (NMDAR shown here, and probably applicable to other antigens) have moderate brain inflammatory infiltrates along with plasma cells (brown cells in **I**), deposits of IgG (diffuse brown staining in **J**), and microglial proliferation (inset in **J**), without evidence of predominant T-cell-mediated neuronal loss (not shown). All human tissue sections (**G–J**) were obtained from hippocampus. (From J Dalmau: Antibody mediated encephalitis. *N Engl J Med* 378:840, 2018. Copyright © 2018 Massachusetts Medical Society. Reprinted with permission.)

inversion recovery [FLAIR] sequences); (3) *brainstem encephalitis*, resulting in eye movement disorders (nystagmus, opsoclonus, supranuclear or nuclear paresis), cranial nerve paresis, dysarthria, dysphagia, unsteady gait, and central autonomic dysfunction; (4) *cerebellar gait and limb ataxia*; (5) *myelitis*, which may cause lower or upper motor

neuron symptoms, myoclonus, muscle rigidity, spasms, sensory deficits, and sphincter dysfunction; and (6) *autonomic dysfunction* as a result of involvement of the neuraxis at multiple levels, including hypothalamus, brainstem, and autonomic nerves (see Paraneoplastic Peripheral Neuropathies, below). Cardiac arrhythmias, postural

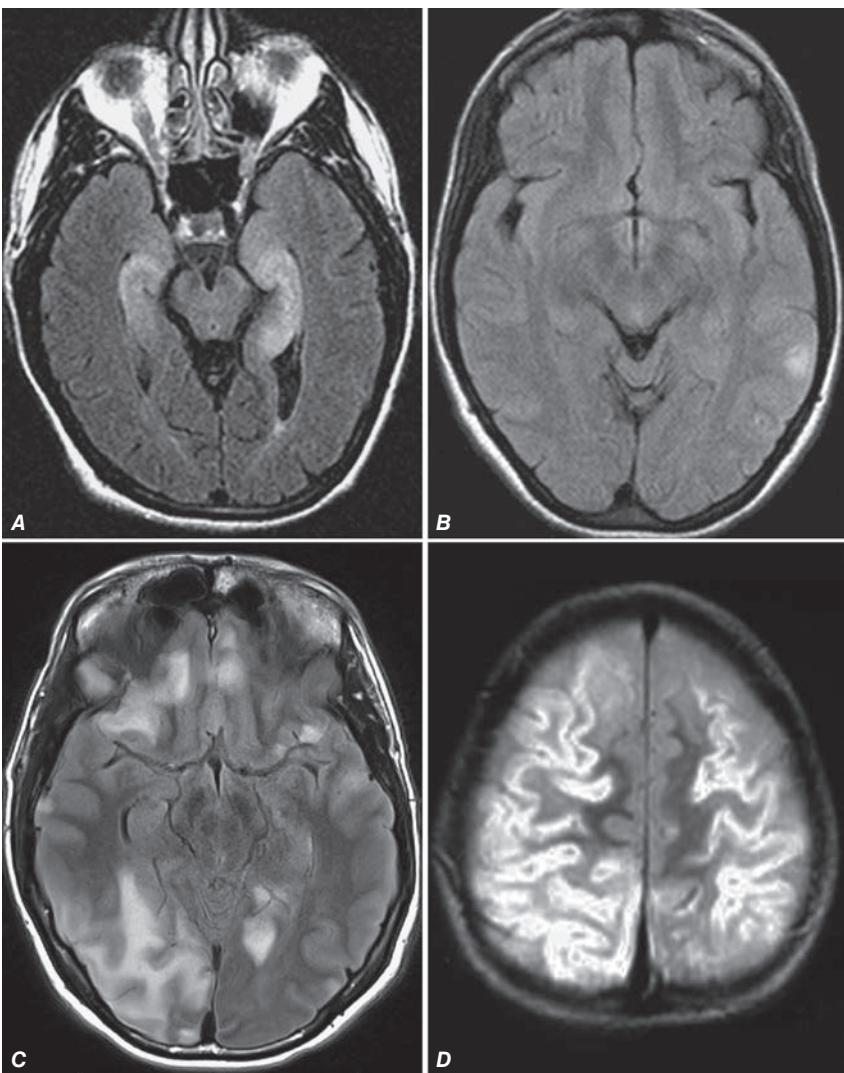


FIGURE 94-2 Brain MRI findings in paraneoplastic and autoimmune encephalitis. Representative MRI studies of patients with several types of autoimmune encephalitides. **A.** Limbic encephalitis (LE) may result from several different immune responses (Hu, Ma2, AMPAR, GABA_AR, LGI1, Caspr2) and typically manifests with unilateral or bilateral medial temporal lobe increased FLAIR signal. **B.** Anti-NMDAR encephalitis often occurs with normal MRI findings or mild FLAIR signal abnormalities. **C.** In contrast, anti-GABA_AR encephalitis usually occurs with multiple cortical-subcortical increased FLAIR signal changes. **D.** Cortical encephalitis can occur in patients with myelin oligodendrocyte glycoprotein (MOG) antibodies, as shown in this T2-weighted MRI image from a 3-year-old boy who presented with extensive cortical abnormalities with mild enhancement (not shown here) suggesting cortical necrosis. (Panels A-C from J Dalmau: Antibody mediated encephalitis. *N Engl J Med* 378:840, 2018. Copyright © 2018 Massachusetts Medical Society. Reprinted with permission. Panel D from T Armangue: Associations of paediatric demyelinating and encephalitic syndromes with myelin oligodendrocyte glycoprotein antibodies: A multicentre observational study. *Lancet Neurol* 19:234, 2020.)

hypotension, and central hypoventilation can be the cause of death in patients with encephalomyelitis.

Paraneoplastic encephalomyelitis and focal encephalitis are usually associated with SCLC, but many other cancers have been implicated. Patients with SCLC and these syndromes usually have Hu antibodies in serum and CSF. CRMP5 antibodies occur less frequently; some of these patients may develop chorea, uveitis, or optic neuritis. Antibodies to Ma proteins are associated with limbic, hypothalamic, and brainstem encephalitis and occasionally with cerebellar symptoms; some patients develop hypersomnia, cataplexy, and severe hypokinesia. MRI abnormalities are frequent, including those described with limbic encephalitis and variable involvement of the hypothalamus, basal ganglia, or upper brainstem. Kelch-like protein 11 antibodies are predominantly associated with brainstem encephalitis and seminomas, germ cell tumors, and teratomas. Amphiphysin antibodies usually are associated with paraneoplastic stiff-person syndrome, but in some patients, they can occur with paraneoplastic encephalomyelitis or isolated myelitis.

The oncologic associations of these antibodies are shown in Table 94-2.

Most types of paraneoplastic encephalitis and encephalomyelitis in which the antigens are intracellular respond poorly to treatment. Stabilization of symptoms or partial neurologic improvement may occur, particularly if there is a satisfactory response of the tumor to treatment. Controlled trials of therapy are lacking, but many reports and the opinion of experts suggest that therapies aimed to remove the antibodies against intracellular antigens, such as intravenous immunoglobulin (IVIg) or plasma exchange, usually fail. The main concern should be to treat the tumor and consider immunotherapies aimed at cytotoxic T-cell responses. Approximately 30% of patients with anti-Ma2-associated encephalitis respond to treatment of the tumor (usually a germ cell neoplasm of the testis) and immunotherapy.

■ ENCEPHALITIDES WITH ANTIBODIES TO CELL-SURFACE OR SYNAPTIC PROTEINS (TABLE 94-3)

These disorders are important for four reasons: (1) they can occur with and without tumor association; less frequently, they develop after a viral encephalitis (herpes simplex or Japanese encephalitis); (2) some syndromes predominate in young individuals and children; (3) despite the severity of the symptoms, patients usually respond to treatment of the tumor, if found, and immunotherapy (e.g., glucocorticoids, IVIg, plasma exchange, rituximab, or cyclophosphamide); and (4) for many of these disorders, the antibody pathogenicity has been demonstrated in models using cultures of neurons or passive transfer of patients' antibodies to animals (Fig. 94-3).

Encephalitis with N-methyl-D-aspartate (NMDA) receptor antibodies usually occurs in young women and children, but men and older patients of both sexes can be affected. The disorder has a characteristic pattern of symptom progression that often includes a prodrome resembling a viral process, followed in a few days by the onset of severe psychiatric symptoms, sleep dysfunction (usually insomnia), reduced verbal output, memory loss, seizures, decreased level of consciousness, abnormal movements (orofacial, limb, and trunk dyskinesias, dystonic postures), autonomic instability, and frequent hypoventilation. Monosymptomatic episodes, such as pure psychosis, occur in about 5% of patients. Clinical relapses occur in 12–24% of patients (12% during the first 2 years after initial presentation). Most patients have intrathecal synthesis of antibodies, likely by infiltrating plasma

cells in brain and meninges (Fig. 94-1J). In about 65% of patients, the brain MRI is normal; in the other 35%, it shows FLAIR abnormalities that can affect cortical and subcortical regions, usually mild and transient, and rarely the presence of contrast enhancement (Fig. 94-2B). The syndrome may be misdiagnosed as a viral or idiopathic encephalitis, neuroleptic malignant syndrome, or encephalitis lethargica, and some patients are initially evaluated by psychiatrists with the suspicion of acute psychosis as the presentation of a primary psychiatric disease. The detection of an associated teratoma is dependent on age and gender: 46% of female patients 12 years or older have uni- or bilateral ovarian teratomas, whereas <7% of girls younger than 12 have a teratoma (Fig. 94-4A). In young male patients, the detection of a tumor is rare. Patients older than 45 years are more frequently male; about 20% of these patients have tumors (e.g., cancer of the breast, ovary, or lung). Prompt diagnosis and treatment with immunotherapy (and tumor removal when it applies) improve outcome. Overall, about 85–90% of patients have substantial neurologic improvement or full recovery.

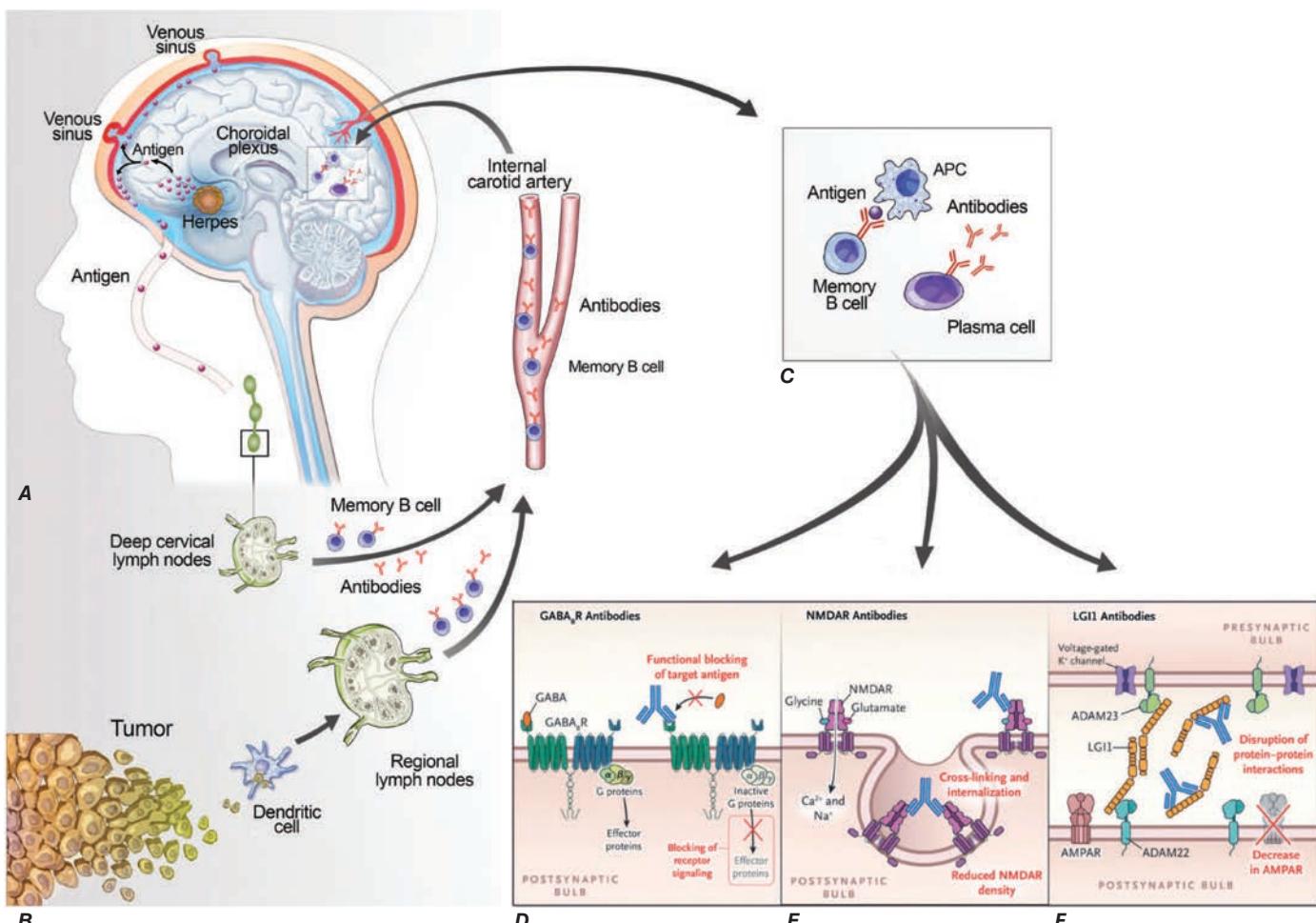


FIGURE 94-3 Proposed mechanisms of disease and functional interactions of autoantibodies with neuronal proteins. The graph shows a multistep process that results in antibody-mediated neuronal dysfunction; some of the steps have been demonstrated in reported studies, whereas others are based on proposed hypotheses. Two well-known triggers of autoimmune encephalitides are represented: herpes simplex encephalitis (**A**) and systemic tumors (**B**); the genetic susceptibility of some autoimmune encephalitides and unknown immunologic triggers are not depicted. It is postulated that antigens released by viral-induced neuronal destruction or apoptotic tumor cells are loaded into antigen-presenting cells (APCs; dendritic cells) and transported to regional lymph nodes. In the lymph nodes, naïve B cells exposed to the processed antigens, with cooperation of CD4+ T cells, become antigen-experienced and differentiate into antibody-producing plasma cells. After entering the brain, memory B cells undergo restimulation, antigen-driven affinity maturation, clonal expansion, and differentiation into antibody-producing plasma cells (**C**). The contribution of systemically produced antibodies to the pool of antibodies present in the brain is unclear and may depend on systemic antibody titers and integrity of the blood-brain barrier. Based on experimental models with cultured neurons, the presence of antibodies in the brain may lead to neuronal dysfunction by different mechanisms, including functional blocking of the target antigen (GABA_AR antibodies; **D**), receptor crosslinking and internalization (NMDAR antibodies; **E**), and disruption of protein-protein interaction, leading to downstream effects on receptors (LGI1 leading to a decrease of Kv1 potassium channels and AMPAR; **F**). These mechanisms are influenced by the type of antibodies; for example, whereas IgG1 antibodies frequently crosslink and internalize the target antigen, IgG4 antibodies are less effective at crosslinking the target and more often alter protein-protein interactions. (Panels D-F J Dalmau: Antibody mediated encephalitis. *N Engl J Med* 378:840, 2018. Copyright © 2018 Massachusetts Medical Society. Reprinted with permission.)

Deficits of attention, memory, and executive functions may recover slowly over many months, sometimes a few years.

Approximately 25% of patients with herpes simplex encephalitis develop a form of autoimmune encephalitis that usually is associated with abnormal movements (choreoathetosis after herpes simplex encephalitis) in children and with cognitive and psychiatric symptoms in adults. This disorder develops a few weeks after the viral infection has resolved, is associated with new synthesis of antibodies against the NMDA receptor and other neuronal cell surface proteins, and is usually less responsive to immunotherapy than anti-NMDA receptor encephalitis (idiopathic or teratoma-associated).

Encephalitis with α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptor antibodies affects middle-aged women, who develop acute limbic dysfunction or, less frequently, prominent psychiatric symptoms; 70% of patients have an underlying tumor in the lung, breast, or thymus (Fig. 94-4B). In about 50% of cases, the brain MRI shows typical features of limbic encephalitis (similar to Fig. 94-2A). Neurologic relapses may occur; these also respond to immunotherapy and are not necessarily associated with tumor recurrence.

Encephalitis with GluK2 antibodies can affect children and adults and is associated with rapidly progressive encephalopathy with cerebellar ataxia or cerebellitis. Symptoms of encephalopathy may include impairment of memory and level of consciousness and motor alterations such as dyskinesias, choreoathetosis, bradykinesia, and spastic paraparesis. Some patients develop intracranial hypertension. In one patient, the symptoms were associated with teratoma.

Encephalitis with γ-aminobutyric acid type A (GABA_A) receptor antibodies may affect children and adults and is associated with prominent seizures and status epilepticus often requiring a pharmacologically induced coma. In approximately 80% of patients, the brain MRI shows multifocal, asynchronous, cortical-subcortical T2/FLAIR abnormalities predominantly involving temporal and frontal lobes, but also basal ganglia and other regions (Fig. 94-2C). Most patients do not have an underlying tumor, but some, usually of Japanese ethnicity, may have thymoma.

Encephalitis with GABA_B receptor antibodies is usually associated with limbic encephalitis and seizures. In >50% of cases, the MRI shows increased medial temporal lobe FLAIR changes characteristic of limbic encephalitis (similar to Fig. 94-2A). In rare instances, patients

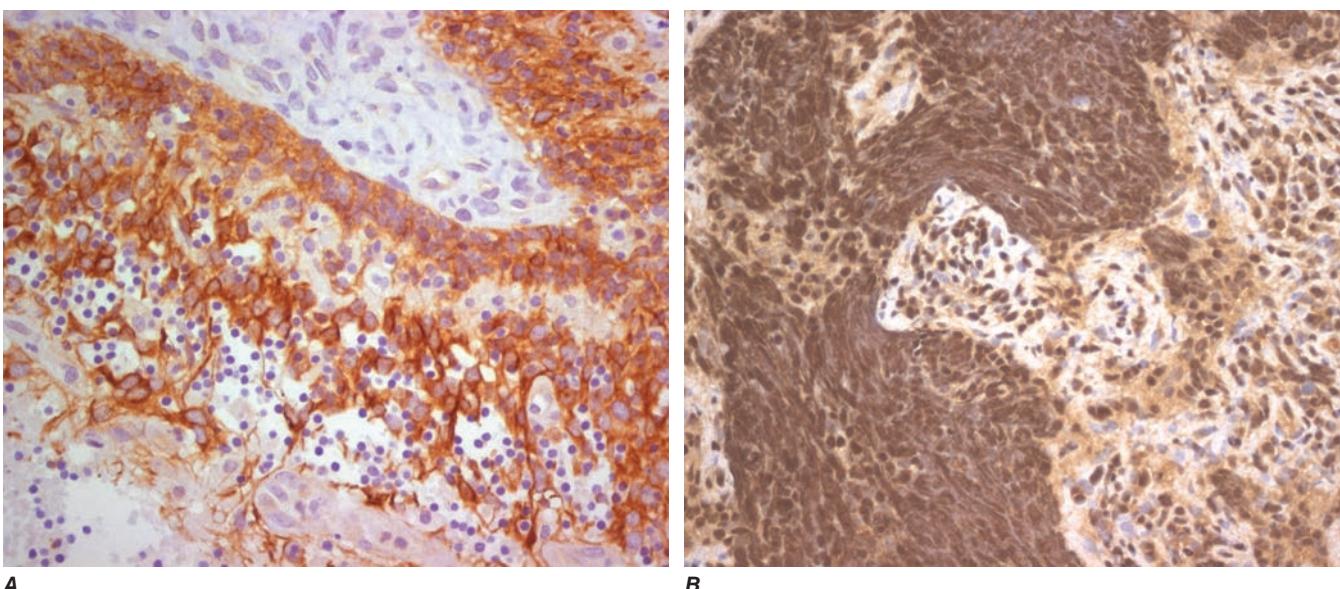


FIGURE 94-4 Immunopathological studies in tumors of patients with autoimmune encephalitis. **A.** Neurons and neuronal processes (brown cells; stained with MAP2) in the teratoma of a patient with anti-NMDA receptor encephalitis; these neurons express NMDA receptors (not shown). **B.** Lung cancer from a patient with anti- α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptor encephalitis showing expression of AMPA receptors by the neoplastic cells (brown cells). (Panel B from M Lai et al: AMPA receptor antibodies in limbic encephalitis alter synaptic receptor location. Ann Neurol 65:424, 2009.)

develop cerebellar symptoms and opsoclonus. Fifty percent of patients have SCLC or a neuroendocrine tumor of the lung. Patients may have additional antibodies to glutamic acid decarboxylase (GAD), which are of unclear significance. Other antibodies to nonneuronal proteins are often found in these patients as well as in patients with AMPA receptor antibodies, indicating a general tendency to autoimmunity.

Encephalitis with glycine receptor (GlyR) antibodies usually manifests with a syndrome characterized by progressive encephalomyelitis with rigidity and myoclonus (PERM) or stiff-person spectrum of symptoms. The disease usually occurs in adults and rarely in children. About 20% of adult patients have a concurrent underlying tumor (thymoma, B-cell lymphoma, breast or lung cancer) or past history of cancer (thymoma, breast, Hodgkin lymphoma, melanoma).

Encephalitis with metabotropic glutamate receptor 5 (mGluR5) antibodies is characterized by nonspecific clinical features of encephalitis (confusion, agitation, memory loss, delusions, paranoid ideation, hallucinations, psychosis, or seizures) without distinctive MRI changes and frequent association with Hodgkin's lymphoma (Ophelia syndrome). The encephalitis is highly responsive to immunotherapy and treatment of the tumor.

Encephalitis with antibodies against dopamine-2 receptor has been reported in children with basal ganglia encephalitis manifesting with abnormal movements (coarse tremor, parkinsonism, chorea, oculogyric crises) along with psychiatric features, lethargy, drowsiness, brainstem dysfunction, or ataxia. The disorder is extremely rare and is not associated with cancer.

Encephalitis with leucine-rich glioma-inactivated 1 (LGI1) antibodies predominates in patients older than 50 years (65% male) and frequently presents with short-term memory loss and seizures (limbic encephalopathy), along with hyponatremia and sleep dysfunction. The MRI often shows increased FLAIR signal in one or both medial temporal lobes. In about 40% of patients, these symptoms are preceded by faciobrachial dystonic seizures, which consist of sudden, short-lasting, mainly distal muscle contractions involving the arm, face, or leg. These are unilateral but can independently affect both sides and occur multiple times during the day or night. About 15% of patients present with rapidly progressive cognitive decline, resembling a rapidly progressive dementia. Less than 5% of patients have thymoma. An association with the human leukocyte antigen (HLA) haplotypes DRB1*07:01, DQB1*02:02, DQA1*02:01, and DRB4 has been identified in non-paraneoplastic cases. All symptoms, including faciobrachial dystonic seizures, respond to immunotherapy, although about two-thirds of patients are left with memory or cognitive deficits.

Encephalitis with contactin-associated protein-like 2 (Caspr2) antibodies predominates in patients older than 50 years and is associated with a form of encephalitis with three or more of the following core symptoms: encephalopathy, cerebellar symptoms, peripheral nervous system hyperexcitability, dysautonomia, insomnia, neuropathic pain, and weight loss. Patients with Morvan's syndrome, which includes clinical features of encephalitis (confusion, hallucinations, prominent sleep dysfunction, or "agrypnia excitata"), autonomic alterations, and peripheral nerve hyperexcitability or neuromyotonia, usually have Caspr2 antibodies. About 20% of patients with Caspr2 antibody-associated syndromes have thymoma; this percentage is higher (~40%) in patients with Morvan's syndrome. An association of Caspr2 antibody-associated syndromes with HLA DRB1*11:01 has been reported.

Encephalitis with dipeptidyl-peptidase-like protein-6 (DPPX) antibodies is usually preceded or develops concurrently with diarrhea, other gastrointestinal symptoms, and substantial loss of weight that often suggest the presence of a gastrointestinal disease. Neurologic symptoms include agitation, hallucinations, paranoid delusions, and features of CNS hyperexcitability such as tremor, myoclonus, nystagmus, seizures, or hyperekplexia. Some patients develop a clinical picture similar to progressive encephalomyelitis with rigidity and myoclonus. The few patients reported with an associated tumor all had B-cell neoplasms.

Encephalitis with antibodies against neurexin 3 alpha does not have distinctive clinical features; the experience is limited, and the disorder does not appear to be associated with cancer.

Anti-IgLON5 disease is a chronic or subacute encephalopathy that characteristically is associated with rapid eye movement (REM) and non-REM (NREM) parasomnia that may be preceded or accompanied by bulbar symptoms, gait abnormalities, movement disorders (chorea, distal myoclonus, tremor, dystonia, or spasms), oculomotor dysfunction, and, in less than half of cases, cognitive decline. The median age of the patients is in the early 60s, and men and women are equally affected. The sleep disorder is characterized by abnormal sleep initiation with undifferentiated NREM sleep associated with frequent vocalizations and quasi-purposeful movements. Examination of the CSF and MRI is unrevealing or demonstrates minor changes of unclear clinical relevance. It is not associated with cancer but shows a strong association with HLA-DRB1*10:01 and HLA-DQB1*05:01. The response to immunotherapy is poor. Neuropathologic studies often show a neuronal tauopathy predominantly involving the hypothalamus and tegmentum of the brainstem.

With the exception of patients with anti-IgLON5 disease, who rarely respond to treatment, most patients with autoimmune or

paraneoplastic encephalopathies associated with antibodies against cell-surface or synaptic proteins respond to immunotherapy and treatment of the tumor (if appropriate). Although there are no specific standardized treatment protocols, the most frequent approach consists of progressive escalation of immunotherapy using first a combination of glucocorticoids, IVIg, and plasma exchange, and then, if there is no response, rituximab or cyclophosphamide.

Encephalitis with myelin oligodendrocyte glycoprotein (MOG) antibodies can present with a clinical picture suggestive of autoimmune encephalitis related to neuronal antibodies. Most patients with MOG antibody-associated syndromes are children and young adults who present with optic neuritis, myelitis, or acute disseminated encephalomyelitis (ADEM). About 85% of patients with these syndromes respond to immunotherapy, although relapses occur in about 30% of cases. Besides these syndromes, there is a small group of adults and children that present with unilateral or bilateral cortical encephalitis, and their response to treatment is variable. In children, two phenotypes of poor prognosis include ADEM-like relapses progressing to leukodystrophy-like features and extensive cortical encephalitis evolving to atrophy (Fig. 94-2D). In general, MOG antibody syndromes are not associated with tumors.

■ PARANEOPLASTIC CEREBELLAR DEGENERATION

This disorder is often preceded by a prodrome that may include dizziness, oscillopsia, blurry or double vision, nausea, and vomiting. A few days or weeks later, patients develop dysarthria, gait and limb ataxia, and variable dysphagia. The examination usually shows downbeating nystagmus and, rarely, opsoclonus. Brainstem dysfunction, upgoing toes, or a mild neuropathy may occur. Early in the course, MRI studies are usually normal; later, the MRI reveals cerebellar atrophy. The disorder results from extensive degeneration of Purkinje cells, with variable involvement of other cerebellar cortical neurons, deep cerebellar nuclei, and spinocerebellar tracts. The tumors more frequently involved are SCLC, cancer of the breast and ovary, and Hodgkin's lymphoma.

Anti-Yo (PCA1) antibodies in patients with breast or gynecologic cancers typically are associated with prominent or pure cerebellar degeneration. A variable degree of cerebellar dysfunction can be associated with virtually any of the antibodies and PND of the CNS shown in Table 94-2. A number of single case reports have described neurologic improvement after tumor removal, plasma exchange, IVIg, cyclophosphamide, rituximab, or glucocorticoids. However, most patients with paraneoplastic cerebellar degeneration and any of the antibodies shown in Table 94-2 do not improve with treatment.

A cerebellar syndrome can also occur with antibodies against cell-surface or synaptic proteins, including P/Q-type voltage-gated calcium channels (VGCC), Tr (DNER), mGluR2, or Sez6l2 (Table 94-3). The frequency and type of tumor association vary with the type of antibody. The cerebellar syndrome of patients with mGluR1 antibodies is highly responsive to treatment of the tumor and immunotherapy, whereas the syndrome of patients with Tr or VGCC antibodies is less treatment responsive. The experience with mGluR2 and Sez6l2 is limited to a few patients, but mGluR2 antibody-associated cerebellar symptoms seem to be highly responsive to treatment. Patients with GluK2 antibodies can present with cerebellitis and posterior fossa edema with compression of the 4th ventricle; the syndrome is potentially treatable with immunotherapy.

■ PARANEOPLASTIC OPSOCLONUS-MYOCLOMUS SYNDROME

Opsoclonus is a disorder of eye movement characterized by involuntary, chaotic saccades that occur in all directions of gaze; it is frequently associated with myoclonus and ataxia. Opsoclonus-myoclonus may be cancer-related or idiopathic. When the cause is paraneoplastic, the tumors involved are usually cancer of the lung and breast in adults, neuroblastoma in children, and ovarian teratoma in adolescents and young women. The pathologic substrate of opsoclonus-myoclonus is unclear, but studies suggest that disinhibition of the fastigial nucleus of the cerebellum is involved. Most patients do not have antineuronal antibodies. A small subset of patients with ataxia, opsoclonus,

and other eye-movement disorders develop Ri antibodies; these patients may also develop muscle rigidity, laryngeal spasms, and autonomic dysfunction. The tumors most frequently involved in anti-Ri-associated syndromes are breast, ovarian, and lung cancers. If the tumor is not successfully treated, the syndrome in adults often progresses to encephalopathy, coma, and death. In addition to treating the tumor, symptoms may respond to immunotherapy (glucocorticoids, plasma exchange, and/or IVIg).

At least 50% of children with opsoclonus-myoclonus have an underlying neuroblastoma. Hypotonia, ataxia, behavioral changes, and irritability are frequent accompanying symptoms. Neurologic symptoms often improve with treatment of the tumor and glucocorticoids, adrenocorticotrophic hormone (ACTH), plasma exchange, IVIg, rituximab, or cyclophosphamide. Many patients are left with psychomotor retardation and behavioral and sleep problems.

■ PARANEOPLASTIC SYNDROMES OF THE SPINAL CORD

The number of reports of paraneoplastic spinal cord syndromes, such as *subacute motor neuronopathy* and *acute necrotizing myelopathy*, has decreased in recent years. This may represent a true decrease in incidence, due to improved and prompt oncologic interventions, or the identification of nonparaneoplastic etiologies. Some patients with cancer or lymphoma develop *upper or lower motor neuron dysfunction* or both, resembling amyotrophic lateral sclerosis. It is unclear whether these disorders have a paraneoplastic etiology or simply coincide with the presence of cancer.

Paraneoplastic myelitis may present with upper or lower motor neuron symptoms, segmental myoclonus, sensory deficits, sphincter dysfunction, and neurogenic pruritus and can be the first manifestation of encephalomyelitis. The spine MRI usually shows longitudinally extensive, symmetric tract or gray matter abnormalities in the spinal cord. It is mainly associated with breast and lung carcinomas and with CRMP5 or amphiphysin antibodies. The prognosis is poor. *Neuromyelitis optica (NMO) with aquaporin 4 antibodies* may occur in rare instances as a paraneoplastic manifestation of a cancer. NMO is discussed in detail in Chap. 445.

■ PARANEOPLASTIC STIFF-PERSON SYNDROME

This disorder is characterized by progressive muscle rigidity, stiffness, and painful spasms triggered by auditory, sensory, or emotional stimuli. Rigidity mainly involves the lower trunk and legs, but it can affect the upper extremities and neck. Sometimes, only one extremity is affected (*stiff-limb syndrome*). Symptoms improve with sleep and general anesthetics. Electrophysiologic studies demonstrate continuous motor unit activity. The associated antibodies target proteins (GAD, amphiphysin) involved in the function of inhibitory synapses using γ -aminobutyric acid (GABA) or glycine as neurotransmitters. The presence of amphiphysin antibodies usually indicates a paraneoplastic etiology related to SCLC and breast cancer. By contrast, GAD antibodies may occur in some cancer patients but are much more frequently present in the nonparaneoplastic disorder. GlyR antibodies may occur in some patients with stiff-person syndrome; these antibodies are more frequently detectable in patients with PERM (Fig. 94-5).

Optimal treatment of stiff-person syndrome requires therapy of the underlying tumor, glucocorticoids, and symptomatic use of drugs that enhance GABAergic transmission (diazepam, baclofen, sodium valproate, tiagabine, vigabatrin). IVIg and plasma exchange are transiently effective in some patients, and there are reports of responses to rituximab in patients who did not respond to other treatments.

■ PARANEOPLASTIC SENSORY NEURONOPATHY OR DORSAL ROOT GANGLIONOPATHY

This syndrome is characterized by sensory deficits that may be symmetric or asymmetric, painful dysesthesias, radicular pain, and decreased or absent reflexes. All modalities of sensation and any part of the body including face and trunk can be involved. Special senses such as taste and hearing can also be affected. Electrophysiologic studies show decreased or absent sensory nerve potentials with normal or

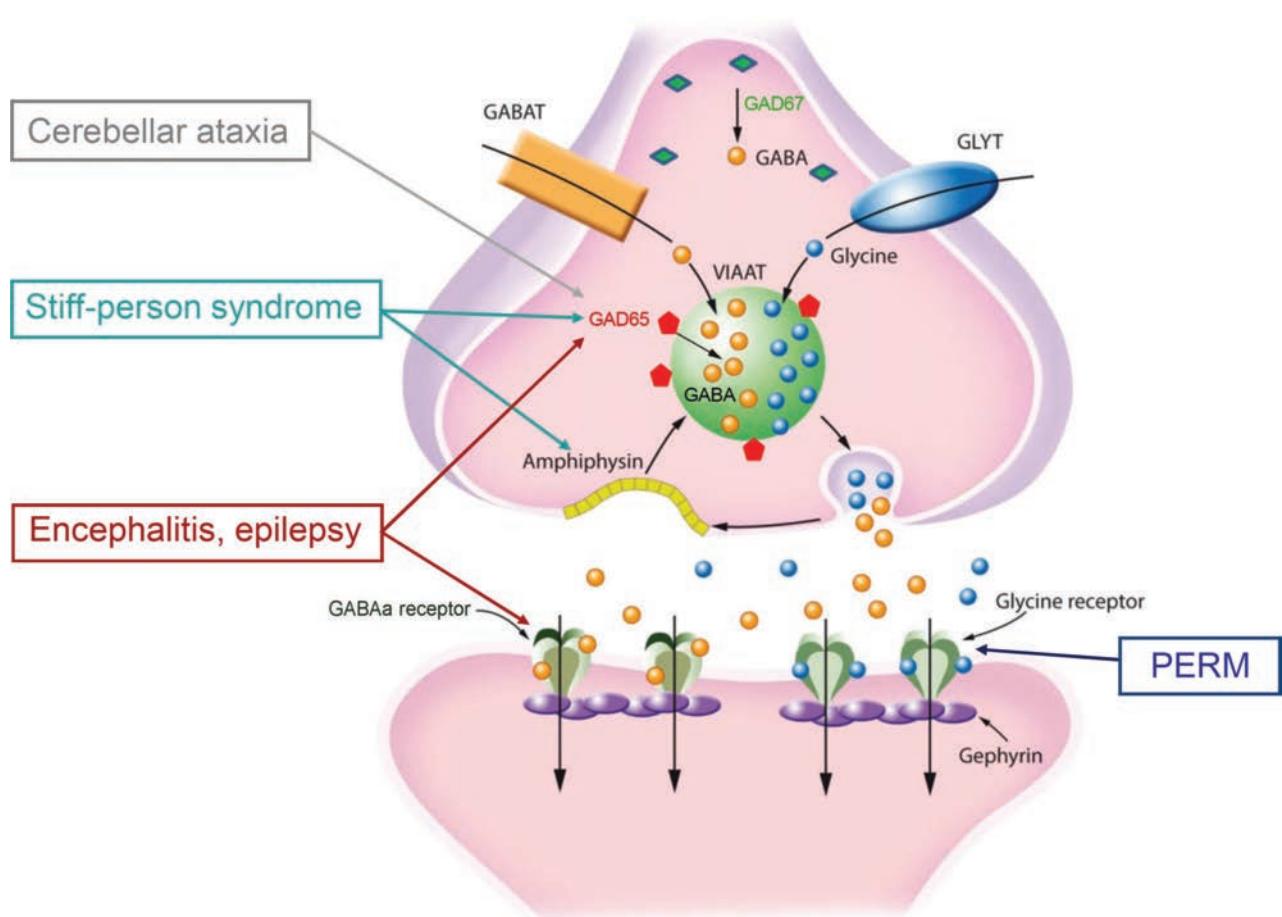


FIGURE 94-5 Schematic representation of an inhibitory synapse showing the main autoimmune targets (GAD, amphiphysin, GABA receptor, and glycine receptor) and the corresponding neurologic disorders. GAD antibodies predominantly occur in stiff-person syndrome (SPS), cerebellar ataxia, and epilepsy, sometimes in the setting of encephalitis. Amphiphysin antibodies are markers of paraneoplastic SPS and breast cancer, GlyR antibodies often associate with progressive encephalomyelitis with rigidity and myoclonus (PERM), and GABA_A receptor antibodies occur in a form of autoimmune encephalitis that is frequently associated with refractory seizures and status epilepticus. (Modified from F Graus et al: *Nat Rev Neurol* 16:353, 2020.)

near-normal motor conduction velocities. Symptoms result from an inflammatory, likely immune-mediated, process that targets the dorsal root ganglia, causing neuronal loss and secondary degeneration of the posterior columns of the spinal cord. The dorsal and, less frequently, the anterior nerve roots and peripheral nerves may also be involved. This disorder often precedes or is associated with encephalomyelitis and autonomic dysfunction and has the same immunologic and oncologic associations (Hu antibodies, SCLC).

As with anti-Hu-associated encephalomyelitis, the therapeutic approach focuses on prompt treatment of the tumor and cytotoxic T-cell-mediated mechanisms. Glucocorticoids occasionally produce clinical stabilization or improvement. The benefit of IVIg and plasma exchange is not proven.

■ PARANEOPLASTIC PERIPHERAL NEUROPATHIES

These disorders may develop any time during the course of the neoplastic disease. Neuropathies occurring at late stages of cancer or lymphoma usually cause mild to moderate sensorimotor deficits due to axonal degeneration of unclear etiology. These neuropathies are often masked by concurrent neurotoxicity from chemotherapy and other cancer therapies. In contrast, the neuropathies that develop in the early stages of cancer frequently show a rapid progression, sometimes with a relapsing and remitting course, and evidence of inflammatory infiltrates and axonal loss or demyelination. If demyelinating features predominate (Chaps. 446 and 447), IVIg, plasma exchange, or glucocorticoids may improve symptoms. Occasionally, CRMP5 antibodies are present; detection of anti-Hu suggests concurrent dorsal root ganglionitis.

Guillain-Barré syndrome (Chap. 447) and brachial plexitis (Chap. 446) have occasionally been reported in patients with

Hodgkin's lymphoma, but there is no clear evidence of a paraneoplastic association.

Diseases associated with monoclonal gammopathies such as multiple myeloma, osteosclerotic myeloma, cryoglobulinemia, amyloidosis, Waldenström's macroglobulinemia, or POEMS (polyneuropathy, organomegaly, endocrinopathy, M-protein spike, and skin manifestations) syndrome, among others, may cause neuropathy by a variety of mechanisms, including compression of roots and plexuses by metastasis to vertebral bodies and pelvis, by deposits of amyloid in peripheral nerves, or through a direct interaction of the abnormal immunoglobulin with peripheral nerve antigens. In other patients, the mechanisms underlying the neuropathy remain unknown and paraneoplastic immune-mediated mechanisms have not been ruled out. Neuropathies more often occur with IgM gammopathies followed by IgG and IgA. The phenotype of the neuropathy and likelihood of improvement with successful treatment of the gammopathy are dependent on the underlying hematologic disorder (Chap. 447).

Vasculitis of the nerve and muscle causes a painful symmetric or asymmetric distal axonal sensorimotor neuropathy with variable proximal weakness. It predominantly affects elderly men and is associated with an elevated erythrocyte sedimentation rate and increased CSF protein concentration. SCLC and lymphoma are the primary tumors involved. Glucocorticoids and cyclophosphamide often result in neurologic improvement.

Peripheral nerve hyperexcitability (neuromyotonia, or Isaacs' syndrome) is characterized by spontaneous and continuous muscle fiber activity of peripheral nerve origin. Clinical features include cramps, muscle twitching (fasciculations or myokymia), stiffness, delayed muscle relaxation (pseudomyotonia), and spontaneous or evoked

carpal or pedal spasms. The involved muscles may be hypertrophic, and some patients develop paresthesias and hyperhidrosis. The electromyogram (EMG) shows fibrillations; fasciculations; and doublet, triplet, or multiplet single-unit (myokymic) discharges that have a high intraburst frequency. Some patients have Caspr2 antibodies usually in the context of Morvan's syndrome, but most patients with isolated neuromyotonia are antibody negative. The disorder often occurs without cancer; if paraneoplastic, benign and malignant thymomas and SCLC are the usual tumors. Phenytoin, carbamazepine, and plasma exchange improve symptoms.

Paraneoplastic autonomic neuropathy usually develops as a component of other disorders, such as LEMS and encephalomyelitis. It may rarely occur as a pure or predominantly autonomic neuropathy with cholinergic or adrenergic dysfunction at the pre- or postganglionic levels. Patients can develop several life-threatening complications, such as gastrointestinal paresis with pseudo-obstruction, cardiac dysrhythmias, and postural hypotension. Other clinical features include abnormal pupillary responses, dry mouth, anhidrosis, erectile dysfunction, and problems in sphincter control. The disorder occurs in association with several tumors, including SCLC, cancer of the pancreas or testis, carcinoid tumors, and lymphoma. Because autonomic symptoms can be the presenting feature of encephalomyelitis, serum Hu and CRMP5 antibodies should be sought. Antibodies to ganglionic (α 3-type) neuronal acetylcholine receptors are the cause of autoimmune autonomic ganglionopathy, a disorder that frequently occurs without cancer association (Chap. 440).

LAMBERT-EATON MYASTHENIC SYNDROME

LEMS is discussed in Chap. 448.

MYASTHENIA GRAVIS

Myasthenia gravis is discussed in Chap. 448.

POLYMYOSITIS-DERMATOMYOSITIS

Polymyositis and dermatomyositis are discussed in detail in Chap. 365.

IMMUNE-MEDIATED NECROTIZING MYOPATHY

Patients with this syndrome develop myalgias and rapid progression of weakness involving the extremities, neck, pharyngeal, respiratory, and sometimes cardiac muscles. Serum muscle enzymes are elevated, and muscle biopsy shows extensive necrosis with minimal or absent inflammation and sometimes deposits of complement. The disorder may occur without cancer association (sometimes as a result of statin exposure, connective tissue disease, or HIV) or with cancer association. Patients with antibodies against 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR) and seronegative patients are more likely to have an underlying cancer than those with antibodies against signal recognition particle. No specific type of cancer has been found to be predominant. Successful treatment of the tumor and aggressive immunotherapy (steroids, IVIg, and steroid-sparing immunosuppressants) may lead to complete or substantial recovery. Immune-mediated necrotizing myopathy is discussed in Chap. 365.

PARANEOPLASTIC VISUAL SYNDROMES

This group of disorders involves the retina and, less frequently, the uvea and optic nerves. The term *cancer-associated retinopathy* is used to describe paraneoplastic cone and rod dysfunction characterized by photosensitivity, progressive loss of vision and color perception, central or ring scotomas, night blindness, and attenuation of photopic and scotopic responses in the electroretinogram (ERG). The most commonly associated tumor is SCLC. Melanoma-associated retinopathy affects patients with metastatic cutaneous melanoma. Patients develop acute onset of night blindness and shimmering, flickering, or pulsating photopsias that often progress to visual loss. The ERG shows reduced b-waves with normal dark adapted a-waves. Paraneoplastic optic neuritis and uveitis can develop in association with encephalomyelitis. Patients with paraneoplastic uveitis and optic neuritis may harbor CRMP5 antibodies.

Some paraneoplastic retinopathies are associated with serum antibodies that specifically react with the subset of retinal cells undergoing degeneration, supporting an immune-mediated pathogenesis (Table 94-2). Paraneoplastic retinopathies rarely show substantial improvement after treatment of the tumor and immunotherapy; however, stabilization of symptoms and partial responses to a variety of immunotherapies (glucocorticoids, plasma exchange, IVIg, rituximab, or alemtuzumab) have been reported.

FURTHER READING

- ANTOINE JC, CAMDESSANCHÉ JP: Paraneoplastic neuropathies. *Curr Opin Neurol* 30:513, 2017.
- ARMANGUE T et al: Associations of paediatric demyelinating and encephalitic syndromes with myelin oligodendrocyte glycoprotein antibodies: A multicentre observational study. *Lancet Neurol* 19:234, 2020.
- BINKS S et al: Distinct HLA associations of LGI1 and Caspr2-antibody diseases. *Brain* 141:2263, 2018.
- DALMAU J et al: Autoantibodies to synaptic receptors and neuronal cell-surface proteins in autoimmune diseases of the central nervous system. *Physiol Rev* 97:839, 2017.
- DALMAU J, GRAUS F: Antibody-mediated encephalitis. *N Engl J Med* 378:840, 2018.
- DARNELL RB, POSNER J: *Paraneoplastic Syndromes*. New York, Oxford University Press, 2011.
- DIDELOT A, HONNORAT J: Paraneoplastic disorders of the central and peripheral nervous systems. *Handb Clin Neurol* 121:1159, 2014.
- FLANAGAN EP et al: Autoimmune myelopathies. *Continuum* 17:776, 2011.
- GRAUS F et al: Clinical approach to diagnosis of autoimmune encephalitis. *Lancet Neurol* 15:391, 2016.
- VAN SONDEREN A et al: The value of LGI1, Caspr2, and voltage-gated potassium channel antibodies in encephalitis. *Nat Rev Neurol* 13:290, 2017.

95

Cancer Survivorship and the Long-Term Impact of Cancer and Its Treatment

Mark Roschewski, Dan L. Longo

The impact of cancer extends well past initial diagnosis. Patients are significantly affected by cancer and treatment-related toxicities often extending beyond the initial treatment period. Adult survivors of childhood, adolescent, and young adult cancer face special health consequences of cancer treatment related to premature physiologic aging and frailty. More than 40% of these patients will experience a severe, disabling, or life-threatening condition or die of a chronic condition. Long-term effects include toxicities that emerge during therapy and continue beyond treatment, while late effects include toxicities that may not emerge for months or years after treatment. Significant improvements in cancer treatments have enabled more people to survive once-deadly diseases, leading to more cancer survivors subjected to the potential long-term impact of cancer treatment (Table 95-1 lists potential long-term and late effects of cancer therapy by organ system). The direct causality of emerging treatments may not be immediately evident, and pharmacovigilance remains critical after treatments first become approved.

TABLE 95-1 Organ Systems at Risk for Long-Term and Late Effects of Cancer Treatment

ORGAN SYSTEM	LONG-TERM EFFECTS	LATE EFFECTS
Cardiovascular	Congestive heart failure	Congestive heart failure
	Arrhythmias	Arrhythmias
	Pericardial disease	Coronary artery disease
	Myocarditis	Peripheral vascular disease
	Hypertension	Cardiac valvular disease
Pulmonary	Pneumonitis	Pulmonary fibrosis Radiation recall pneumonitis
Immunologic	Opportunistic infections	Second malignancies
	Autoimmune disease	Myelodysplasia Recurrent infections Autoimmune disease
Endocrine	Fractures	Gonadal dysfunction/infertility
	Hypopituitarism	Premature ovarian insufficiency
	Avascular necrosis	Sarcopenia
	Diabetes insipidus	Diabetes Osteoporosis Thyroid disorders
Neurologic	Peripheral sensorimotor neuropathy Myopathy Hearing loss	Cognitive impairment
Gastrointestinal	Malabsorption Colitis Chronic liver disease	Chronic diarrhea Small bowel obstruction Gastrointestinal stricture
Genitourinary	Chronic renal failure	Hemorrhagic cystitis Ureteral stricture
Psychological	Anxiety Depression	Mood disorders Posttraumatic stress disorder Sexual dysfunction Substance abuse disorders Financial hardship Psychosocial dysfunction

In the United States, the number of cancer survivors may increase from 17 million to nearly 26 million by the year 2040, and the number of patients who survive at least 5 years after initial diagnosis is expected to increase by 35% over the next decade. Improvements in cancer treatments for children and adolescents have led to modern 5-year survival rates of approximately 80% or greater. Despite the magnitude of the growing problem, the core issues related to cancer survivorship remain understudied and the research is often concentrated in highly prevalent cancers. Most studies are observational and descriptive with fewer studies focused on the prevention and treatment of complications. Cancer survivorship remains an area that is ripe for further discovery; a deeper understanding of the biological basis and/or the influence of genetics on host susceptibility and the long-term effects of cancer therapy is needed.

Our primary understanding of the long-term impact of cancer treatment originated from survivors of childhood malignancies who are often cured after treatment with a combination of chemotherapy, radiation, and surgery. Treatment paradigms for cancer continuously evolve, however, and newer treatments including targeted agents and immunotherapy have introduced unique long-term effects, particularly when these agents are administered indefinitely (Table 95-2 lists the potential long-term effects of specific cancer treatments). Improvements in the tolerability of cancer therapy and in supportive care have allowed a greater number of patients with comorbid conditions and advanced age to receive treatment, which has increased the rate of

chronic morbidity. Approximately 60% of current cancer survivors are older than age 65. Indeed, the cause-specific mortality following initial treatment of Hodgkin lymphoma includes nearly 40% of deaths attributed to nonlymphoma causes. Indeed, more patients diagnosed with Hodgkin lymphoma die from treatment-related late toxicity than from Hodgkin lymphoma. Individuals aged 60–74 have a disproportionate excess of deaths related to heart disease, lung disease, infections, and adverse effects of drugs. The complexities of cancer survivorship and the importance of longitudinal care of cancer patients are recognized as vital components to comprehensive cancer care, and model systems have been specifically developed for this purpose. Still, the primary goal of cancer therapy remains long-term disease control, and the treating physician must maintain proper perspective when considering these relative risks. The fear of long-term complications should not prevent the application of effective cancer treatment, particularly when delivered with curative intent. In a sense, managing long-term effects is a privilege only afforded those fortunate enough to overcome the initial threat to life represented by the cancer diagnosis.

CARDIOVASCULAR DYSFUNCTION

The excess risks of cardiovascular disease after anthracycline chemotherapy and radiation that involves the mediastinum are well characterized and include arrhythmias, cardiac ischemia, congestive heart failure (CHF), pericardial disease, myocarditis, and peripheral vascular disease. It is estimated that one in eight childhood cancer survivors will experience a life-threatening cardiovascular event within 30 years after initial exposure, and cardiovascular disease is the most common cause of noncancer death in this population. Newer targeted agents and immunotherapy have introduced additional cardiovascular risks that extend into older populations as well. A new discipline of cardio-oncology has been developed to better characterize individuals at high risk for treatment-related cardiac toxicities, develop surveillance strategies, and improve the management of long-term effects.

Radiation therapy that includes the heart can cause interstitial myocardial fibrosis, acute and chronic pericarditis, valvular disease, and accelerated premature atherosclerotic coronary artery disease. Repeated or high (>6000 cGy) radiation doses are associated with greater risk, as is concomitant cardiotoxic cancer chemotherapy exposure. Symptoms of acute pericarditis peak about 9 months after treatment and include dyspnea, chest pain, and fever. Chronic constrictive pericarditis may develop 5–10 years following radiation therapy. Cardiac valvular disease includes aortic insufficiency from fibrosis or papillary muscle dysfunction resulting in mitral regurgitation. Extensive radiation fields are associated with accelerated coronary artery disease and peripheral vascular disease and a markedly increased risk of fatal myocardial infarction or thromboembolic stroke. Three-dimensional conformal techniques and newer particles including proton beams may more precisely target the tumor and spare normal tissue, but these are not widely available and the degree to which they will decrease long-term cardiovascular effects is unknown. In recognition of the risks of radiation, careful planning procedures are performed before treatment designed to limit the field of radiation to the greatest extent possible.

The myocardial toxicity of anthracyclines is dose-dependent and is associated with the pathognomonic finding of myofibrillar dropout on endomyocardial biopsy. Anthracycline cardiotoxicity occurs through a root mechanism of chemical free radical damage. Fe³⁺-doxorubicin complexes damage DNA, nuclear and cytoplasmic membranes, and mitochondria. These cardiotoxic effects may also be mediated by topoisomerase IIB. Approximately 5% of patients receiving >450–550 mg/m² total dose of doxorubicin will develop CHF, but it can also develop at substantially lower doses in some patients. Anthracycline-related CHF is often irreversible and carries a high mortality rate, making prevention crucial. Genome-wide association studies have identified multiple genetic polymorphisms associated with a higher risk of cardiotoxicity, but our ability to risk-stratify is limited. The risk of cardiac failure appears to be related to the route of administration, and regimens that use continuous infusion of doxorubicin or liposomally encapsulated doxorubicin are associated with less cardiotoxicity. Baseline assessment of cardiac function with multigated acquisition scan (MUGA) or

TABLE 95-2 Long-Term Effects of Cancer Treatment by Type of Therapy

THERAPY TYPE	LONG-TERM EFFECT
Radiation	Second malignancies Coronary artery disease Pericardial disease Peripheral vascular disease Cardiac valvular disease Neurocognitive impairment Hypopituitarism and infertility Hypothyroidism and reduced bone density Gastrointestinal stricture Hepatic venoocclusive disease
Chemotherapy and Hormonal Agents	
Anthracyclines, trastuzumab, cyclophosphamide	Congestive heart failure
Bleomycin, oxaliplatin, 5-fluorouracil	Pulmonary fibrosis
Nitrosoureas, methotrexate, fludarabine, brentuximab	Pneumonitis
Alkylating agents, anthracyclines, tamoxifen, bendamustine, platinum agents	Second malignancies/myelodysplasia
Bendamustine, alkylating agents, anthracyclines	Immune dysfunction and recurrent infections
Alkylating agents	Infertility
Cyclophosphamide, Ifosfamide	Hemorrhagic cystitis
Platinum agents	Renal tubular dysfunction
Vinca alkaloids, taxanes, platinum agents	Neuropathy
Cytarabine	Ataxia
Aromatase inhibitors	Vasomotor symptoms
Antiandrogens	Sexual dysfunction
Immunotherapy Agents	
Immune checkpoint inhibitors	Autoimmune conditions/autoimmune hepatitis
Immune checkpoint inhibitors	Pericarditis and fulminant myocarditis
CAR-T therapy	Congestive heart failure
CAR-T therapy	Arrhythmias
CAR-T therapy, bi-specific monoclonal antibodies, immune checkpoint inhibitors	Peripheral neuropathy
CAR-T therapy, bi-specific monoclonal antibodies	B-cell aplasia
Targeted Agents	
Immunomodulatory agents	Second cancers/leukemia
Proteasome inhibitors	Peripheral neuropathy
Anti-CD20 agents, immunomodulatory agents	Neutropenia
BTK inhibitors, ALK inhibitors	Atrial and ventricular arrhythmias
PI3K inhibitors, CDK inhibitors	Hepatitis, colitis
Gemtuzumab	Sinusoidal obstruction syndrome
EGFR inhibitors, anti-VEGF agents, BCR-ABL inhibitors, MEK inhibitors, proteasome inhibitors	Congestive heart failure
BCR-ABL inhibitors	Pleural effusions, pancreatitis
BCR-ABL inhibitors, PI3K inhibitors, BTK inhibitors, immunomodulatory agents, EGFR inhibitors	Chronic diarrhea
BCR-ABL inhibitors	Impaired growth and stature
Anti-VEGF agents, BCR-ABL inhibitors	Thyroid dysfunction
PI3K inhibitors, mTOR inhibitors, BRAF inhibitors	Hyperglycemia
FLT3 inhibitors, anti-VEGF agents, PI3K inhibitors	Systemic hypertension
Anti-VEGF agents, BCR-ABL inhibitors	Pulmonary hypertension

Abbreviation: CAR-T, chimeric antigen receptor T cell.

transthoracic echocardiograms is commonly performed, and a patient who develops symptoms suggestive of CHF should be tested immediately while therapy is held. Periodic surveillance testing during therapy is often done in asymptomatic patients with preexisting risk factors.

Trastuzumab is a monoclonal antibody targeting human epidermal growth factor receptor 2 (HER2) that is also associated with CHF. It is used in combination with chemotherapy as both adjuvant therapy and as treatment of metastatic breast cancer, and it is sometimes combined

with anthracyclines, which is believed to result in additive or possibly synergistic toxicity. In contrast to anthracyclines, cardiotoxicity is not dose related, is usually reversible, is not associated with pathologic changes on cardiac myofibrils, and has a different biochemical mechanism inhibiting intrinsic cardiac repair mechanisms. Monitoring for cardiac toxicity is typically performed every three or four doses using functional cardiac testing, and treatment is interrupted when ejection fractions significantly decline from baseline. Other potentially

cardiotoxic chemotherapy agents include phosphoramide mustards (cyclophosphamide) at high doses and ifosfamide.

Small-molecule inhibitors, including tyrosine kinase inhibitors, are novel classes of molecularly targeted anticancer agents that have become routinely applicable across a variety of malignancies. Although the overall tolerability of these drugs is often better than chemotherapy, they are frequently administered indefinitely, which introduces new notions of cumulative long-term effects. These agents also carry risks of cardiovascular toxicities including CHF, atrial and ventricular arrhythmias, prolongation of the QT interval, and pulmonary and systemic hypertension. New anticancer agents often become available for use on accelerated approvals before a full understanding of the long-term toxicity profile is known. Two illustrative examples of this are lapatinib and ponatinib, which both had their approvals updated with black box cardiovascular toxicity warnings a median of 4 years after initial drug approval. Other small-molecule inhibitors that have been associated with CHF include bosutinib, dasatinib, nilotinib, pazopanib, axitinib, trametinib, sunitinib, carfilzomib, and sorafenib. Systemic hypertension is commonly associated with agents targeting vascular endothelial growth factor or its receptors (e.g., bevacizumab, cabozantinib, lenvatinib, nintedanib), ponatinib (an Abl inhibitor), and trametinib (a MEK inhibitor), whereas dasatinib (an Abl inhibitor) has a well-documented association with pulmonary hypertension. Ibrutinib (a BTK inhibitor) has been associated with atrial fibrillation as well as ventricular arrhythmias. As more of these small-molecule inhibitors become approved for use and their indications broaden, additional monitoring for long-term and late effects should be incorporated into the routine surveillance of oncologists, cardiologists, and primary care providers.

Immunotherapy agents have also emerged as effective anticancer therapies that have substantially improved clinical outcomes in a variety of cancers. Immune checkpoint inhibitors have been associated with a number of important cardiovascular toxicities including pericardial disease, vasculitis, and fulminant myocarditis. The mechanism of these toxicities is T cell mediated, and the toxicities often respond to early institution of glucocorticoids, but they can be severe or fatal if not recognized promptly. Combinations of multiple immune checkpoint inhibitors increase the risk of these immune-related toxicities, and no clear pattern of which patients are most susceptible has emerged. Chimeric antigen receptor T-cell (CAR-T) therapies are associated with cytokine release syndromes (CRS), which can be severe and associated with arrhythmias or decompensated heart failure. Interleukin 6 receptor blockers may decrease these risks, and supportive care guidelines recommend early institution of these agents in severe CRS.

The management of treatment-associated cardiovascular disease is essentially the same as for cardiac disease unrelated to cancer treatment. Discontinuation of the offending agent is the first step. Diuretics, fluid and sodium restriction, and antiarrhythmic agents are often useful for acute symptoms. Afterload reduction with angiotensin-converting enzyme inhibitors or β -adrenergic blockers may improve systolic function over time, and digitalis may improve symptoms. Routine screening for asymptomatic systolic dysfunction is currently recommended for survivors at high risk for cardiomyopathy including those with an anthracycline exposure $\geq 250 \text{ mg/m}^2$, $\geq 35 \text{ Gy}$ of chest radiation, or combined therapy with both anthracyclines and radiation. Echocardiography is the recommended screening modality, and surveillance should begin no later than 2 years after exposure and should be repeated a minimum of every 5 years thereafter.

PULMONARY DYSFUNCTION

Radiation-induced lung injury presents in early phases as acute pneumonitis at 4 weeks following treatment, but it can evolve into pulmonary fibrosis in late phases. Risk factors for radiation pneumonitis include advanced age, smoking, poor performance status, preexisting compromised pulmonary function, and radiation volume and dose. It occurs most commonly in patients with lung cancer, mediastinal lymphoma, and breast cancer, and the incidence is decreasing due to advances in radiation delivery techniques. The dose “threshold” is thought to be in the range of 5–20 Gy. Hypoxemia and dyspnea on

exertion are characteristic, and the severity of symptoms may be out of proportion to the lung volume irradiated. Fine, high-pitched “Velcro rales” may be an accompanying physical finding, and fever, cough, and pleuritic chest pain are common symptoms. The diffusion capacity of the lungs for carbon dioxide (D_{LCO}) is the most sensitive measure of pulmonary functional impairment, and ground-glass infiltrates often correspond with relatively sharp edges to the irradiated volume, although the pneumonitis may progress beyond the field and even occasionally involve the contralateral unirradiated lung. The mechanism of lung injury is a direct effect of radiation that leads to increased capillary permeability and pulmonary edema. Damage to type I and II pneumocytes leads to surfactant loss and the transudation of serum proteins into the alveoli. Cytokines including tumor necrosis factor α are released from the damaged lung cells and attract inflammatory cells to the alveoli and interstitial space. The late phases of injury are caused by reactive oxygen species that stimulate collagen production and lead to fibrosis but do not occur in all cases. Transforming growth factor β (TGF- β) is particularly important in stimulating collagen synthesis and may represent a therapeutic target to prevent pulmonary fibrosis.

Bleomycin generates activated free radical oxygen species and causes pneumonitis associated with a radiographic or interstitial ground-glass appearance diffusely throughout both lungs, often worse in the lower lobes. A nonproductive cough with or without fever may be an early sign. This toxicity is dose-related and dose-limiting. The D_{LCO} is a sensitive measure of toxicity and recovery, and a baseline value is generally obtained for future comparison before administering bleomycin therapy. Doses are reduced or stopped if the baseline D_{LCO} falls 25% or more. Additive or synergistic risk factors include age, prior lung disease, and concomitant use of other chemotherapy, lung irradiation, and high concentrations of inspired oxygen. Other chemotherapeutic agents notable for pulmonary toxicity include mitomycin, nitrosoureas, doxorubicin with radiation, gemcitabine combined with weekly docetaxel, methotrexate, and fludarabine. High-dose alkylating agents, cyclophosphamide, ifosfamide, and melphalan are frequently used in the hematopoietic stem cell transplant setting, often with whole-body radiation. This therapy may result in severe pulmonary fibrosis and/or pulmonary venoocclusive disease.

Radiation-induced lung injury and chemotherapy-induced pneumonitis are generally glucocorticoid responsive, except in the case of nitrosoureas. Prednisone 1 mg/kg is often used to control acute symptoms and prevent pulmonary dysfunction with a slow taper over 12 weeks. Prolonged glucocorticoid therapy requires gastrointestinal protection with proton pump inhibitors, management of hyperglycemia, heightened infection management, and prevention or treatment of steroid-induced osteoporosis. Antibiotics, bronchodilators, oxygen in only lowest necessary doses, and diuretics may all play an important role in management of pneumonitis, and consultation with a pulmonologist should be routinely undertaken. Relapses can occur after an initial response to glucocorticoids and may respond to agents such as azathioprine or cyclosporine. Amifostine is a free radical scavenger and radioprotective agent that reduces the rate of pneumonitis, but it is associated with severe nausea and hypotension that limit its use. No effective therapy exists for pulmonary fibrosis, and the treatment is primarily supportive with supplemental oxygen. Targeted anti-inflammatory agents are being tested to reduce the incidence of pulmonary fibrosis, but they remain experimental.

Pulmonary toxicity resulting from targeted anticancer agents including small-molecule inhibitors and immunotherapy agents is uncommon but can be potentially life-threatening and often lead to drug discontinuation. Noninfectious pneumonitis is associated with cough, dyspnea, and infiltrates on chest imaging and has been reported to be associated with sunitinib, sorafenib, epidermal growth factor receptor (EGFR) inhibitors (cetuximab, afatinib), crizotinib (ALK inhibitor), phosphoinositide 3-kinase (PI3K) inhibitors (idelalisib, copanlisib), and mammalian target of rapamycin (mTOR) inhibitors (everolimus, temsirolimus). The antibody-drug conjugate brentuximab vedotin may also cause severe pulmonary toxicity when used in combination with other chemotherapy agents, particularly bleomycin. The onset of drug-induced pneumonitis can be rapid, and prompt use

of glucocorticoids is important once infectious causes are excluded. Severe pneumonitis is typically a reason to discontinue the offending drug permanently.

IMMUNE SYSTEM DYSFUNCTION

A significant risk of most anticancer treatment is hematologic toxicity with cumulative effects on the host immune system leading to a higher risk of second malignancies and impaired long-term immune health. Second malignancies in cancer survivors are a major cause of death, and survivors of childhood cancers have a twofold increased risk of solid tumors beyond the age of 40 years compared to the general population. The induction of second malignancies is governed by the complex interplay of age, sex, environmental exposures, genetic susceptibility, and specific cancer treatments. Often, the events that led to the primary cancer remain, and a risk of a second malignancy persists. Patients with a history of lung cancer remain at increased risk of other cancers that are associated with tobacco use including esophageal, head and neck, kidney, and bladder cancers. Patients with breast cancer are at increased risk of breast cancer in the opposite breast. Patients with Hodgkin lymphoma are at risk for other B-cell non-Hodgkin lymphomas. Genetic cancer syndromes (e.g., multiple endocrine neoplasia or Li-Fraumeni, Lynch's, Cowden's, and Gardner's syndromes) are examples of genetically based second malignancies of specific types. Cancer treatment itself does not appear to be responsible for the risk of these secondary malignancies. Genetic disorders that result in DNA repair deficiencies including ataxia-telangiectasia, Bloom's syndrome, and Fanconi's anemia greatly increase the lifetime risk of cancers as well as the risks associated with DNA-damaging agents. Importantly, the risk of treatment-related second malignancies is at least additive and often synergistic with combined chemotherapy and radiation therapy, and hence for such combined-therapy treatment approaches, it is important to establish the necessity of each component in the treatment program. These patients require indefinite surveillance and prophylactic surgery in some cases.

Patients receiving radiation have an increasing and lifelong risk of second malignancies that is 1–2% in the second decade following treatment but increases to >25% after 25 years. The risk of second malignancies from radiation is dose-dependent and often occurs within or near the treatment field. Common radiation-related solid tumors include central nervous system (CNS), breast, lung, thyroid, skin, and bone cancers and sarcomas, which are often aggressive and have a poor prognosis. An example of an organ-, age-, and sex-dependent radiation-induced secondary malignancy is breast cancer, in which the risk is small with radiation in women aged older than 30 years but increases about twentyfold over baseline in women aged younger than 30 years. A 25-year-old woman treated with mantle radiation for Hodgkin lymphoma has a 29% actuarial risk of developing breast cancer by age 55.

Chemotherapy is significantly associated with two fatal second malignancies: acute leukemia and myelodysplastic syndromes. Two types of secondary leukemia have been described; in patients treated with chronic alkylating agents (especially combined with radiation therapy), acute myeloid leukemia is associated with deletions in chromosome 5 or 7 and complex karyotypes and often is preceded by myelodysplasia. The lifetime risk is about 1–5%, is increased by radiation therapy, and increases with age. The incidence of these leukemias peaks at 5–8 years, with risk returning close to baseline at 10 years. The other type of acute myeloid leukemia is related to therapy with topoisomerase inhibitors, is associated with chromosome 11q23 translocations, has an incidence of <1%, generally occurs 1.5–3 years after treatment, and is rarely preceded by myelodysplasia. Both of these acute myeloid leukemias are largely refractory to treatment and have a high mortality. The development of myelodysplastic syndromes is increased following chemotherapy, and these cases are often associated with leukemic progression and a dismal prognosis. A fraction of the population develops clonal hematopoiesis not related to prior cancer treatment, and the percentage increases with age. In such patients, the hematopoietic stem cells carry mutations that are associated with myeloid malignancy despite normal blood counts. It is thought that the presence of these genetic lesions may predispose patients to develop

myeloid malignancies, but evidence is greater that clonal hematopoiesis increases the risk of lymphoma and atherosclerotic heart disease.

Other cytotoxic agents are associated with long-term alterations in immunity beyond the initial treatment period and neutrophil recovery. Bendamustine has significant effects on both B-cell and T-cell subpopulations that can persist for years, and long-term studies of lymphoma patients treated with bendamustine-based regimens show higher rates of death from second malignancies compared to other chemotherapy regimens. Purine analogues like cladribine and pentostatin also produce long-term T-cell suppression. The risk of solid tumors is also increased after chemotherapy, and alkylating agents increase the risk of thyroid, lung, breast, and bladder cancers and sarcomas. Cyclophosphamide increases the risk of both sarcoma and breast cancer in a dose-dependent manner. Other chemotherapy agents, including procarbazine and platinum agents, have been associated with gastrointestinal malignancies. Treatment of breast cancer with tamoxifen for 5 years or longer is associated with a 1–2% risk of endometrial cancer. Surveillance is generally effective at finding these cancers at an early stage. The risk of mortality from tamoxifen-induced endometrial cancer is low compared to the benefit of tamoxifen as adjuvant therapy for breast cancer. Treatment of multiple myeloma with the immunomodulatory agent lenalidomide is associated with a significantly increased risk of second hematologic malignancies including lymphomas and leukemias. These risks are highest after prior use of the alkylating agent melphalan.

Given the high risk of second malignancies in cancer survivors, patients need indefinite surveillance. Guidelines for breast cancer surveillance in survivors exposed to chest radiation recommend that patients be screened with mammograms and/or breast MRI beginning at age 25 years or 8 years after treatment, whichever occurs later. Any organ in the treatment field is susceptible to developing a cancer; e.g., radiation to the chest may increase the risk of gastric or esophageal cancer. Patients exposed to abdominal or pelvic radiation should have annual colonoscopies starting at age 35 years or 10 years after exposure. For patients treated with neck radiation, no formal surveillance is recommended, but fine-needle aspiration should be performed on any palpable thyroid nodules and thyroid-stimulating hormone (TSH) levels should be monitored periodically.

Combination chemotherapy is also associated with impaired immune health and increases the risk of opportunistic infections, autoimmune complications, and impaired host protection from infections. Survivors of lymphoma have elevated risks of developing autoimmune hemolytic anemia, viral or fungal pneumonias, meningitis, or other infections, and these risks remain high decades after treatment. Agents or treatment regimens that result in significant T-cell depletion, including antithymocyte globulin or antibodies targeting cell surface proteins on T cells, increase the risk of Epstein-Barr virus-associated B-cell lymphoproliferative disorders. Discontinuing immunosuppressive therapy, if possible, is often associated with complete disease regression. Anti-CD20 monoclonal antibodies, CAR-T therapy, bi-specific monoclonal antibodies targeting both B cells and T cells, and immune checkpoint inhibitors have all been associated with long-term B-cell aplasia that often requires intravenous immunoglobulin replacement and persistent vigilance for recurrent sinopulmonary infections. Rituximab and immunomodulatory agents have both been associated with late-onset neutropenia that can occur months after exposure to the drug and often requires growth factor support. Given these risks of impaired immune system function, all cancer survivors should undergo annual influenza vaccination and should be considered for pneumococcal vaccination depending on age and immune health status.

REPRODUCTIVE AND ENDOCRINE DYSFUNCTION

Endocrine complications are prevalent in childhood cancer survivors. Nearly half of all survivors will have at least one hormonal disorder in their lifetime, and these most commonly present as late effects. Radiation to the head, neck, or pelvis is associated with the greatest risk of endocrine dysfunction. Testicles and ovaries in prepubertal patients are sensitive to radiation damage in a dose-related fashion;

spermatogenesis is affected by low doses of radiation, and complete azoospermia occurs at 600–700 cGy. Leydig cell dysfunction, in contrast, occurs at <2000 cGy, and hence, endocrine function is lost at much higher radiation doses than spermatogenesis. Erectile dysfunction occurs in up to 80% of men treated with external-beam radiation therapy for prostate cancer. Sildenafil may be useful in reversing erectile dysfunction. Ovarian function damage with radiation is age-related and occurs at doses of 150–500 cGy. Hormone replacement therapy is often contraindicated (as in estrogen receptor-positive breast cancer). Attention must be paid to maintenance of bone mass with calcium and vitamin D supplements and oral bisphosphonates, and bone mass should be monitored using bone density determinations. Paroxetine, clonidine, pregabalin, and other drugs may be useful in symptomatically controlling hot flashes. Long-term survivors of childhood cancer who have received cranial radiation may have altered leptin biology and growth hormone deficiency, leading to obesity and reduced strength, exercise tolerance, and bone density. Radiation therapy to the neck may lead to hypothyroidism, Graves' disease, thyroiditis, and thyroid malignancies. TSH is followed routinely in such patients to prevent hypothyroidism and to suppress persistently elevated levels of TSH, which may cause or drive thyroid cancer. Cranial radiation may also be associated with an array of endocrine abnormalities with disruption of normal pituitary-hypothalamic axis function, and a high index of suspicion needs to be maintained to identify and treat this toxicity. Efforts to eliminate unnecessary radiation such as prophylactic CNS irradiation may decrease some of the late endocrine effects. Patients who have received abdominal radiation should receive annual screens for obesity and diabetes mellitus with height and weight measurements along with a hemoglobin A_{1c} at least every 2 years.

Alkylating agents are the chemotherapy agents associated with the highest rates of male and female infertility, which is directly dependent on age, dose, and duration of treatment. The age at treatment is an important determinant of fertility outcome, with prepubertal patients having the highest tolerance. Ovarian failure is age related, and females who resume menses after treatment are still at increased risk for premature menopause. Males generally have reversible azoospermia during lower intensity alkylator chemotherapy, and long-term infertility is associated with doses of cyclophosphamide >9 g/m² and with high-intensity therapy, such as that used in hematopoietic stem cell transplantation. All patients should be counseled on the potential impact on future reproduction, and timely referral for established interventions such as sperm cryopreservation, oocyte preservation, and embryo preservation should be offered when feasible and appropriate. Assisted reproductive technologies can be helpful to couples with chemotherapy-induced infertility.

Combination chemotherapy can impair bone health, and older patients may be more susceptible to these effects. Due to the combined risk of age-related osteoporosis and the effect of therapy, the risk of fractures in patients over the age of 70 years may be as high as 5–10% within a few years of finishing therapy. The risk of low bone mineral density is highest in certain high-risk groups including survivors of pediatric acute lymphoblastic leukemia and CNS tumors and those who have undergone hematopoietic stem cell transplant.

Immune checkpoint inhibitors that target CTLA-4 and PD-1 have led to serious chronic toxicities including the breaking of self-tolerance and the autoimmune destruction of certain endocrine organs, particularly the thyroid and adenohypophysis (anterior pituitary). Hypophysitis is more commonly reported in association with CTLA-4 inhibitors, whereas thyroid dysfunction is more common with PD-1 inhibitors. Most immune-related toxicities occur within 8–12 weeks of starting treatment, but they can occur at any time during treatment or even after therapy has stopped. Patients with autoimmune thyroiditis or hypophysitis require lifelong hormone replacement, and early recognition is important.

Tyrosine kinase inhibitors such as imatinib have been associated with growth deceleration in children, particularly when treatment is initiated before puberty. The mechanism is postulated to be related to disruptions in growth hormone signaling or inhibition of the insulin-like growth factor 1 (IGFR-1) receptor. Other BCR-ABL inhibitors,

including nilotinib and dasatinib, have been associated with both hyper- and hypothyroidism. Endocrine effects that have been reported to be associated with other tyrosine kinase inhibitors include alterations in bone remodeling, reduced calcium and vitamin D levels, thyroid dysfunction, gonadal dysfunction, adrenal dysfunction, altered glucose metabolism, and secondary hyperparathyroidism. Thyroid function tests should be monitored periodically while patients are on these targeted agents, and replacement hormones and/or vitamins should be prescribed as necessary.

NEUROLOGIC DYSFUNCTION

Neurologic dysfunction from cancer treatment is increasing in both incidence and severity as a result of improved supportive care that enables more aggressive regimens, an expanded number of older patients receiving treatment, extended durations of therapy, and longer periods of cancer survivorship. Direct effects on myelin, glial cells, and neurons have all been implicated, with alterations in cellular cytoskeleton, axonal transport, and cellular metabolism as potential mechanisms. Telomere shortening that occurs with normal aging may be accelerated by radiation. Survivors of CNS tumors are at the greatest risk of late-onset neurocognitive impairment that includes impaired intelligence and slower processing speeds along with deficits in executive function, memory, and attention. These toxicities are reported after treatment with both radiation and chemotherapy in childhood survivors.

Acute radiation CNS toxicity occurs within weeks and is characterized by nausea, drowsiness, hypersomnia, and ataxia, which typically recover over time. Early delayed toxicity occurring weeks to 3 months following therapy is associated with similar symptoms as acute toxicity and is pathologically associated with reversible demyelination. Chronic, late radiation injury occurs 9 months to up to 10 years following therapy, and dysfunction increases over time. Radiation-associated spinal cord injury (myelopathy) is highly dose-dependent and rarely occurs with modern radiation therapy. An early, self-limited form involving electric sensations down the spine on neck flexion (Lhermitte's sign) is seen 6–12 weeks after treatment and generally resolves over weeks. Peripheral nerve toxicity is quite rare owing to relative radiation resistance. Diffuse radiation injury is associated with global CNS neurologic dysfunction and diffuse white matter changes on computed tomography (CT) or MRI. Pathologically, small vessel changes are prominent and focal necrosis is common. Necrotizing encephalopathy is the most severe form of radiation injury and almost always is associated with concurrent use of chemotherapy, notably methotrexate. Prophylactic cranial irradiation in both childhood and adult leukemias/lymphomas has largely been abandoned due to the acute and long-term effects of therapy. Glucocorticoids may be symptomatically useful for acute toxicities but do not alter the course. Psychostimulants such as methylphenidate may improve attention and executive functioning in childhood survivors.

In children and adolescent cancers, younger age, higher cranial irradiation dose, larger brain volumes irradiated, and longer treatment times are associated with worse neurocognitive outcomes. In adult cancers, patients over the age of 60 who receive whole-brain radiation therapy are at high risk for neurocognitive impairment after therapy. Genetic polymorphisms may be associated with an increased risk of neurocognitive problems, and emerging evidence suggests polymorphisms in the folate pathway, oxidative stress genes, and enzymes that regulate both catecholamines and deamination of amines are associated with individual risk.

Vinca alkaloids produce a characteristic "stocking-glove" neuropathy with numbness and tingling advancing to loss of motor function, which is highly dose related. Distal sensorimotor polyneuropathy prominently involves loss of deep tendon reflexes with initially loss of pain and temperature sensation, followed by proprioceptive and vibratory loss. This requires careful patient history and physical examination by experienced oncologists to decide when the drug must be stopped or reduced to prevent permanent damage. Milder toxicity often slowly completely resolves after treatment discontinuation. Vinca alkaloids may sometimes be associated with jaw claudication,

hoarseness, autonomic neuropathy, ileus, cranial nerve palsies, and, in severe cases, encephalopathy, seizures, and coma. Cisplatin is associated with sensorimotor neuropathy and hearing loss, especially at doses $>400 \text{ mg/m}^2$, requiring audiometry in patients with preexisting hearing compromise. Carboplatin is often substituted in such cases given its lesser effect on hearing. Many of the agents that target kinase enzymes in tumor cells and 5-fluorouracil congeners produce dysesthesias and painful hands and feet known as hand-foot syndrome or palmar-plantar erythrodysesthesia. Symptoms usually abate when the agent is stopped. Methotrexate alone may cause acute leukoencephalopathy characterized by somnolence and confusion that is often reversible. Acute toxicity is dose-related, especially at doses $>3 \text{ g/m}^2$, with younger patients being at greater risk. Subacute methotrexate toxicity occurs weeks after therapy and is often ameliorated with glucocorticoid therapy. Chronic methotrexate toxicity (leukoencephalopathy) develops months or years after treatment and is characterized clinically as progressive loss of cognitive function and focal neurologic signs, which are irreversible, promoted by synchronous or metachronous radiation therapy, and more pronounced at a younger age. Neurocognitive decline following chemotherapy alone occurs notably in breast cancer patients receiving adjuvant chemotherapy with anthracyclines, taxanes, or cyclophosphamide; this has been referred to as “chemo brain.” It is clinically associated with impaired memory, learning, attention, and speed of information processing. There is no clear mechanistic explanation for its cause and no clearly effective therapy, although regular exercise is associated with improved symptoms. Most symptoms improve within a year of therapy, but symptoms can persist in 10–20% of patients for extended periods of time.

Newer molecularly targeted agents and immunotherapy have also been associated with neurologic dysfunction and may exacerbate persistent neuropathy from prior therapy. Proteasome inhibitors are associated with neuropathic pain and motor neuropathy that can occur immediately or be delayed in onset. The proposed mechanism is through enhanced oxidative stress on neural cells. Subcutaneous administration of these agents is associated with less peripheral neuropathy than intravenous infusions. Immunotherapy agents such as CAR-T therapies and bi-specific monoclonal antibodies targeting both T cells and B cells are associated with significant acute neurotoxicity including confusion, encephalopathy, seizures, and cerebellar symptoms. These symptoms are hypothesized to be related to the CRS associated with these therapies and often are time limited. However, a minority of patients treated with these agents will have persistent central and neurologic dysfunction for which management is mainly supportive. Glucocorticoids may be useful in the short term, but treatment with interleukin 6 receptor antagonists that are effective for decreasing the severity of CRS is largely ineffective at preventing neurologic toxicities. Progressive multifocal leukoencephalopathy (PML) is a rare but serious brain infection that is caused by the JC virus and has been reported as a rare complication of treatment with rituximab, ibrutinib, and PI3K inhibitors. The inflammatory response to the virus presents as unifocal or multifocal hyperintense lesions involving the subcortical white matter that are best seen on T2/fluid-attenuated inversion recovery (FLAIR) images on MRI. Treatment is supportive and includes removal of the offending agent. Other targeted agents that may be associated with neurologic dysfunction when given for extended durations of treatment include dasatinib, thalidomide, and lenalidomide.

Antibody-drug conjugates (ADCs) are novel therapies in which a monoclonal antibody targeted to a tumor antigen is attached to a potent anticancer agent via a chemical linker. Often, these agents are associated with significant central and peripheral neurotoxicity, as has been described with agents such as brentuximab vedotin and pertuzumab. These neuropathies often emerge during treatment similar to those seen with chemotherapy and are dose-dependent. Early recognition of treatment-emergent neuropathy induced by ADCs mandates dose interruption or modification to less frequent dosing schedules in order for patients to remain on therapy. Immune checkpoint inhibitors have been associated with autoimmune complications and unique neurologic manifestations such as optic neuritis that may be reversible after glucocorticoids.

HEPATIC AND GASTROINTESTINAL DYSFUNCTION

Long-term hepatic damage from standard chemotherapy regimens is rare. Long-term methotrexate or high-dose chemotherapy alone or with radiation therapy, for example, in preparative regimens for bone marrow transplantation, may result in venoocclusive disease of the liver. This potentially lethal complication classically presents with anicteric ascites, elevated alkaline phosphatase, and hepatosplenomegaly. Pathologically, there is venous congestion, epithelial cell proliferation, and hepatocyte atrophy progressing to frank fibrosis. Frequent monitoring of liver function tests during any chemotherapy is necessary to avoid both idiosyncratic and expected toxicities. Certain nucleoside drugs have been associated with hepatic dysfunction; however, this complication is rare in oncology. Hepatic radiation damage depends on dose, volume, fractionation, preexisting liver disease, and synchronous or metachronous chemotherapy. In general, radiation doses to the liver $>1500 \text{ cGy}$ can produce hepatic dysfunction with a steep dose-injury curve. Radiation-induced liver disease closely mimics hepatic venoocclusive disease.

Novel targeted agents including immunotherapy agents have introduced a number of gastrointestinal toxicities that can occur late in the course of treatment including hepatitis, colitis, malabsorption, and chronic diarrhea. Early signs of serious liver injury should lead to discontinuation of the offending agent as the effect does not appear to be dose-related and dose reductions do not reliably reduce further liver injury. The mechanisms for colitis or hepatitis associated with these targeted agents are not completely understood but are hypothesized to be T cell-mediated, and the risk is highest when used in targeted agent combinations. Diarrhea with or without severe colitis can be associated with virtually all targeted agents including PI3K inhibitors, BCR-ABL inhibitors, BTK inhibitors, EGFR inhibitors, MEK inhibitors, CDK inhibitors, and immunomodulatory agents. Even if the diarrhea is not severe, the impact associated with indefinite treatment greatly interferes with the quality of life and often leads to discontinuation of targeted therapy if not managed effectively. Immunomodulatory agents including lenalidomide are associated with late onset of diarrhea that is caused by bile acid malabsorption and often responds to bile acid sequestrants. Immune checkpoint inhibitors are associated with colitis and hepatitis that may be responsive to prompt initiation of glucocorticoids that may require a long taper for resolution.

RENAL AND BLADDER DYSFUNCTION

Cisplatin produces reversible decrements in renal function but may also produce severe irreversible toxicity in the presence of renal disease and may predispose to accentuated damage with subsequent renal insults. Cyclophosphamide and ifosfamide are prodrugs primarily activated in the liver with cleavage products (acrolein) that can produce hemorrhagic cystitis. This can be prevented with the free radical scavenger MESNA (mercaptoethane sulfonate), which is required for ifosfamide administration. Hemorrhagic cystitis caused by these agents may predispose to bladder cancer.

Targeted agents generally do not carry significant acute nephrotoxic risks, but a number of agents, including PI3K inhibitors, anti-VEGF agents, and FLT3 inhibitors, are associated with systemic hypertension, which can lead to late effects or a progressive decline in renal function. Renal dysfunction following immunotherapy is uncommon, but acute interstitial nephritis can occur. Similar to other immune-related toxicities, this acute toxicity requires prompt use of glucocorticoids to avoid long-term effects on renal function.

PSYCHOLOGICAL DYSFUNCTION AND SOCIOECONOMIC IMPACT OF SURVIVORSHIP

The diagnosis and treatment of cancer can introduce long-term and late psychological effects that continue throughout life. Cancer survivors are at increased risk for anxiety, depression, attention problems, and posttraumatic stress syndromes. Many cancer patients experience intrusive or debilitating concerns about cancer recurrence following

successful therapy. In addition, these patients may experience socio-economic stressors that affect employment, insurance, relationships and lead to financial and/or sexual difficulties. Survivors of childhood cancers are less likely to graduate from college or gain full-time employment than their peers and are more likely to engage in risky health behaviors such as substance abuse and excessive alcohol use. The long-term psychosocial effects of treatment are greatest in patients who undergo CNS-directed therapies including radiation and intensive combination chemotherapy regimens. Oncologists should ask about and address these issues explicitly with patients and provide appropriate counseling or support systems. The overall risk of suicidal ideation and suicide is low but is greater in cancer patients and survivors than age-matched controls. Tailored cognitive-behavioral therapy may improve the anxiety and posttraumatic stress associated with cancer survivorship.

CANCER SURVIVORSHIP CARE PLANS

Survivorship starts at the time of diagnosis and continues indefinitely. Many guidelines recommend that every patient be provided with a survivorship care plan unique to their situation, but the evidence that these improve health outcomes is limited and sufficient resources to implement recommendations are often lacking. Focused surveillance plans for late effects are critical for early detection and implementation of interventions but also must include risk stratification to avoid unnecessary surveillance testing that wastes resources and leads to overdiagnosis and/or psychological distress. Survivorship care has traditionally been performed by oncologists, but the scope of the problem mandates that primary care physicians, midlevel providers, and preventive medicine specialists be trained in the follow-up of treated cancer patients in remission. All former cancer patients should undergo surveillance for recurrence and second malignancies and be monitored for long-term effects of treatment; however, as a practical matter, nearly all recurrences are detected because of symptoms. Health promotion and disease prevention with age- and sex-specific routine screening tests (e.g., colonoscopy, Pap smears, mammography, human papillomavirus vaccination, dual-energy x-ray absorptiometry scans) should be a focus of survivorship care along with psychosocial well-being. Annual mammography should start no later than 10 years after breast radiation. Patients receiving radiation fields encompassing thyroid tissue should have regular thyroid examinations and TSH testing. Localized pain or palpable abnormality in a previously radiated field should prompt radiographic evaluation. Patients treated with alkylating agents or topoisomerase inhibitors should have a complete blood count every 6–12 months, and cytopenias, abnormal cells on peripheral smear, or macrocytosis should be evaluated with bone marrow biopsy and aspirate and include cytogenetics, flow cytometry, or fluorescence in situ hybridization (FISH) studies as appropriate.

As the population of cancer survivors increases and patients live longer, cancer survivorship has become increasingly important, and the Institute of Medicine and National Research Council have published a monograph entitled *From Cancer Patient to Cancer Survivor: Lost in Transition*. The monograph proposes a plan that would inform clinicians caring for cancer survivors of the complete details of patients' previous treatments, complications thereof, signs and symptoms of late effects, and recommended screening and follow-up procedures.

OUTLOOK

Survivorship care is one of the most challenging problems facing oncologists today. The challenge is to develop cancer treatments that utilize the most effective combination of surgery, chemotherapy, radiation, targeted agents, or immunotherapy that is required to cure disease or effect long-term disease control with the least amount of toxicity. As cancer treatments continue to improve, the need for cancer care increases due to more cancer survivors with increasing life expectancy. Clearly, much work remains to elucidate the pathophysiology of cancer treatment-related effects and identification of patient characteristics associated with an increased vulnerability to adverse effects. Clinical management strategies focused on the clinical management of acute toxicities and prevention of long-term effects after therapy are

necessary. Finally, research initiatives should recognize that as treatment paradigms continue to evolve, the nature and biologic basis for toxicities will change. Advances in genomic medicine may add depth to our understanding of toxicities and allow for more personalized surveillance strategies. Longitudinal monitoring of the health of cancer survivors is required since the incidence of late effects of treatment does not appear to plateau over time.

ACKNOWLEDGMENT

We would like to acknowledge the contribution of Carl E. Freter who coauthored a previous version of this chapter; material from his chapter was retained in this version.

FURTHER READING

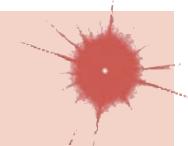
- ARMENIAN SH et al: Cardiovascular disease in survivors of childhood cancer: Insights into epidemiology, pathophysiology, and prevention. *J Clin Oncol* 36:2135, 2018.
- BRINKMAN TM et al: Psychological symptoms, social outcomes, socioeconomic attainment, and health behaviors among survivors of childhood cancer: Current state of the literature. *J Clin Oncol* 36:2190, 2018.
- CHAO C et al: Chronic comorbidities among survivors of adolescent and young adult cancer. *J Clin Oncol* 38:3161, 2020.
- CHEMAITILLY W et al: Endocrine late effects in childhood cancer survivors. *J Clin Oncol* 36:2153, 2018.
- CHOW EJ et al: New agents, emerging late effects, and the development of precision survivorship. *J Clin Oncol* 36:2231, 2018.
- ROWLAND JH et al: Survivorship science at the NIH: Lessons learned from grants funded in fiscal year 2016. *J Natl Cancer Inst* 111:109, 2019.
- SHAPIRO CL: Cancer survivorship. *N Engl J Med* 379:2438, 2018.
- SHAPIRO CL et al: ReCAP: ASCO core curriculum for cancer survivorship education. *J Oncol Pract* 12:e08, 2016.
- SHREE T et al: Impaired immune health in survivors of diffuse large B-cell lymphoma. *J Clin Oncol* 38:1664, 2020.
- TURCOTTE LM et al: Risk, risk factors, and surveillance of subsequent malignant neoplasms in survivors of childhood cancer: A review. *J Clin Oncol* 36:2145, 2018.

Section 2 Hematopoietic Disorders

96

Hematopoietic Stem Cells

David T. Scadden, Dan L. Longo



All of the cell types in the peripheral blood and some cells in every tissue of the body are derived from hematopoietic (*hemo*: blood; *poiesis*: creation) stem cells. If the hematopoietic stem cell is damaged and can no longer function (e.g., due to a nuclear accident), a person would survive 2–4 weeks in the absence of extraordinary support measures. With the clinical use of hematopoietic stem cells, tens of thousands of lives are saved each year (Chap. 114). Stem cells produce hundreds of billions of blood cells daily from a stem cell pool that is estimated to be only 100,000. How stem cells do this, how they persist for many decades despite the production demands, and how they may be better used in clinical care are important issues in medicine.

The study of blood cell production has become a paradigm for how other tissues may be organized and regulated. Basic research in hematopoiesis includes defining stepwise molecular changes accompanying functional changes in maturing cells, aggregating cells into functional subgroups, and demonstrating hematopoietic stem cell

regulation by a specialized microenvironment; these concepts are worked out in hematopoiesis and offer models for other tissues. Moreover, these concepts may not be restricted to normal tissue function but extend to malignancy. Stem cells are rare cells among a heterogeneous population of cell types, and their behavior is assessed mainly in experimental animal models involving reconstitution of hematopoiesis. Thus, much of what we know about stem cells is imprecise and based on inferences from genetically manipulated animals.

CARDINAL FUNCTIONS OF HEMATOPOIETIC STEM CELLS

All stem cell types have two cardinal functions: self-renewal and differentiation (Fig. 96-1). Stem cells exist to generate, maintain, and repair tissues. They function successfully if they can replace a wide variety of shorter-lived mature cells over prolonged periods. The process of self-renewal (see below) assures that a stem cell population can be sustained over time. Without self-renewal, the stem cell pool would become exhausted and tissue maintenance would not be possible. The process of differentiation leads to production of the effectors of tissue function: mature cells. Without proper differentiation, the integrity of tissue function would be compromised and organ failure or neoplasia would ensue.

In the blood, mature cells have variable average life spans, ranging from hours for mature neutrophils to a few months for red blood cells to many years for memory lymphocytes. However, the stem cell pool is the central, durable source of all blood and immune cells, maintaining a capacity to produce a broad range of cells from a single cell source, yet keeping itself vigorous over decades of life. As an individual stem cell divides, it has the capacity to accomplish one of three division outcomes: two stem cells, two cells destined for differentiation, or one stem cell and one differentiating cell. The former two outcomes are the result of symmetric cell division, whereas the latter indicates a different outcome for the two daughter cells—an event termed asymmetric cell division. The relative balance for these types of outcomes may change during development and under particular kinds of demands on the stem cell pool.

DEVELOPMENTAL BIOLOGY OF HEMATOPOIETIC STEM CELLS

During development, blood cells are produced at different sites. Initially, the yolk sac provides oxygen-carrying red blood cells and many of the macrophage-like cells that are resident in tissues: cells like microglia in the brain. The placenta and several sites of intraembryonic blood cell production then become involved in sequential order. These

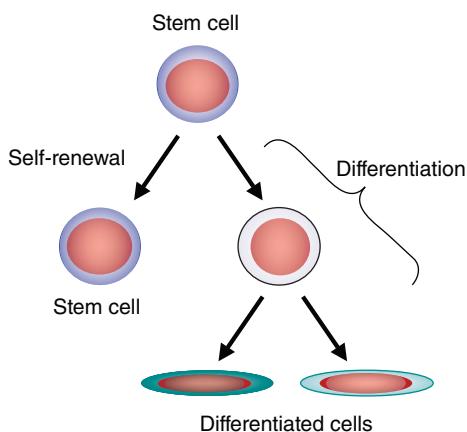


FIGURE 96-1 Signature characteristics of the stem cell. Stem cells have two essential features: the capacity to differentiate into a variety of mature cell types and the capacity for self-renewal. Intrinsic factors associated with self-renewal include expression of *Bmi-1*, *Gfi-1*, *PTEN*, *STAT5*, *Tel/Atv6*, *p21*, *p18*, *MCL-1*, *Mel-18*, *RAE28*, and *HoxB4*. Extrinsic signals for self-renewal include *Notch*, *Wnt*, *SHH*, *angiogenin*, and *Tie2/Ang-1*. Based mainly on murine studies, hematopoietic stem cells express the following cell surface molecules: *CD34*, *Thy-1* (*CD90*), *c-Kit receptor* (*CD117*), *CD133*, *CD164*, and *c-Mpl* (*CD110*, also known as the thrombopoietin receptor).

move from the genital ridge at a site where the aorta, gonadal tissue, and mesonephros are emerging to the fetal liver and then, in the second trimester, to the bone marrow and spleen. As the location of stem cells changes, the cells they produce also change. The yolk sac provides red cells expressing embryonic hemoglobins and tissue-resident macrophages. Intraembryonic sites of hematopoiesis generate stem cells, red cells, platelets, and the circulating cells of innate immunity. The production of the cells of adaptive immunity occurs then as well but becomes robust as the thymus forms and the bone marrow is colonized in the second trimester. Stem cell proliferation remains high, even in the bone marrow, until shortly after birth, when it appears to dramatically decline. The cells in the bone marrow are thought to arrive by the bloodborne transit of cells from the fetal liver after calcification of the long bones has begun. The presence of stem cells in the circulation is not unique to a time window in development, however, as hematopoietic stem cells circulate throughout life. The time that stem cells spend freely circulating appears to be brief (measured in minutes in the mouse), but the stem cells that do circulate are functional and can be used for transplantation. The number of stem cells that circulate can be increased in a number of ways to facilitate harvest and transfer to the same or a different host.

MOBILITY OF HEMATOPOIETIC STEM CELLS

Cells entering and exiting the bone marrow do so through a series of molecular interactions. Circulating stem cells (through CD162 and CD44) engage the lectins (carbohydrate binding proteins) P- and E-selectin on the endothelial surface to slow the movement of the cells to a rolling phenotype. Stem cell integrins are then activated and accomplish firm adhesion between the stem cell and vessel wall, with a particularly important role for stem cell VCAM-1 engaging endothelial VLA-4. The chemokine CXCL12 (SDF1) interacting with stem cell CXCR4 receptors and ionic calcium interacting with the calcium-sensing receptor appear to be important in the process of stem cells getting from the circulation to where they engraft in the bone marrow. This is particularly true in the developmental move from fetal liver to bone marrow.

In the adult, the role for CXCR4 is in retention of stem cells in the bone marrow as well as getting them there. Interrupting that retention process through specific molecular blockers of the CXCR4/CXCL12 interaction, cleavage of CXCL12, or downregulation of the CXCR4 receptor can result in the release of stem cells into the circulation. This process is an increasingly important aspect of recovering stem cells for therapeutic use as it has permitted the harvesting process to be done by leukapheresis rather than bone marrow punctures in the operating room. Granulocyte colony-stimulating factor and plerixafor, a macrocyclic compound that can block CXCR4, are both used clinically to mobilize marrow hematopoietic stem cells for transplant. Refining our knowledge of how stem cells get into and out of the bone marrow may improve our ability to obtain stem cells and make them more efficient at finding their way to the specific sites for blood cell production, the so-called stem cell niche.

HEMATOPOIETIC STEM CELL MICROENVIRONMENT

The concept of a specialized microenvironment, or stem cell niche, was first proposed to explain why cells derived from the bone marrow of one animal could be used in transplantation and again be found in the bone marrow of the recipient. This niche is more than just a housing site for stem cells, however. It is an anatomic location where regulatory signals are provided that allow the stem cells to thrive, to expand if needed, and to provide varying amounts of descendant daughter cells. In addition, unregulated growth of stem cells may be problematic based on their undifferentiated state and self-renewal capacity. Thus, the niche must also regulate the number of stem cells produced. In this manner, the niche has the dual function of serving as a site of nurture but imposing limits for stem cells: in effect, acting as both a nutritive and constraining home.

The niche for blood stem cells changes with each of the sites of blood production during development, but for most of human life, it

is located in the bone marrow. Within the bone marrow, the perivascular space particularly in regions of trabecular bone serves as a niche. The mesenchymal and endothelial cells of the marrow microvessels produce kit ligand and CXCL12, both known to be important for hematopoietic stem cells. Other cell types, such as sympathetic neurons, nonmyelinating Schwann cells, macrophages, megakaryocytes, osteoclasts, and osteoblasts, have been shown to regulate stem cells, some by direct and others by indirect effects. Extracellular matrix proteins like osteopontin and heparan sulfates also affect stem cell function. The endosteal region appears to be particularly important for transplanted cells, in part because many of the mesenchymal cells and sinusoidal blood vessels of the central marrow are disrupted by the conditioning regimens used to prepare a patient for transplantation. The functioning of the niche as a supportive context for stem cells is of obvious importance for maintaining hematopoiesis and in transplantation. An active area of study involves determining whether the niche is altered in disease as experimental models have shown that mutations in niche cells can lead to myeloid malignancies. It logically follows that targeting of niche functions is a potential therapeutic strategy for both malignant and normal hematopoiesis.

■ EXCESS CAPACITY OF HEMATOPOIETIC STEM CELLS

In the absence of disease, one never runs out of hematopoietic stem cells. Indeed, serial transplantation studies in mice suggest that sufficient stem cells are present to reconstitute several animals in succession, with each animal having normal blood cell production. The fact that allogeneic stem cell transplant recipients also never run out of blood cells in their life span, which can extend for decades, argues that even the limiting numbers of stem cells provided to them are sufficient. How stem cells respond to different conditions to increase or decrease their mature cell production remains poorly understood. Clearly, negative feedback mechanisms affect the level of production of most of the cells, leading to the normal tightly regulated blood cell counts. However, many of the regulatory mechanisms that govern production of more mature progenitor cells do not apply or apply differently to stem cells. Similarly, most of the molecules shown to be able to change the size of the stem cell pool have little effect on more mature blood cells. For example, the growth factor erythropoietin, which stimulates red blood cell production from precursor cells, has no effect on stem cells. Similarly, granulocyte colony-stimulating factor drives the rapid proliferation of granulocyte precursors but has little or no effect on the cell cycling of stem cells. Rather, it changes the location of stem cells by indirect means, altering molecules such as CXCL12 that tether stem cells to their niche. Molecules shown to be important for altering the proliferation, self-renewal, or survival of stem cells, such as cyclin-dependent kinase inhibitors, transcription factors like Bmi-1, microRNA-processing enzymes like Dicer, or even metabolic regulators like pyruvate kinase isoforms, have little or different effects on progenitor cells. Hematopoietic stem cells have governing mechanisms that are distinct from the cells they generate.

■ HEMATOPOIETIC STEM CELL DIFFERENTIATION

Hematopoietic stem cells sit at the base of a branching hierarchy of cells culminating in the many mature cell types that compose the blood and immune system (Fig. 96-2). The maturation steps leading to terminally differentiated and functional blood cells take place both as a consequence of intrinsic changes in gene expression and niche-directed and cytokine-directed changes in the cells. Our knowledge of the details remains incomplete. As stem cells mature to progenitors, precursors, and, finally, mature effector cells, they undergo a series of functional changes. These include the obvious acquisition of functions defining mature blood cells, such as phagocytic capacity or hemoglobin synthesis. They also include the progressive loss of plasticity (i.e., the ability to become other cell types). For example, the myeloid progenitor can make all cells in the myeloid series but none in the lymphoid series. As common myeloid progenitors mature, they become precursors for either monocytes and granulocytes or erythrocytes and megakaryocytes, but not both. Some amount of reversibility of this

process may exist early in the differentiation cascade, but that is lost beyond a distinct stage in normal physiologic conditions. With genetic interventions, however, blood cells, like other somatic cells, can be reprogrammed to become a variety of cell types.

As cells differentiate, they may also lose proliferative capacity (Fig. 96-3). Mature granulocytes are incapable of proliferation and only increase in number by increased production from precursors. The exceptions to the rule are some tissue-resident macrophages, which appear capable of proliferation, and lymphoid cells. Lymphoid cells retain the capacity to proliferate but have linked their proliferation to the recognition of particular proteins or peptides by specific antigen receptors on their surface. Like many tissues with short-lived mature cells such as the skin and intestine, blood cell proliferation is largely accomplished by a more immature progenitor population. In general, cells within the highly proliferative progenitor cell compartment are also relatively short-lived, making their way through the differentiation process in a defined molecular program involving the sequential activation of particular sets of genes. For any particular cell type, the differentiation program is difficult to speed up. The time it takes for hematopoietic progenitors to become mature cells is ~10–14 days in humans, evident clinically by the interval between cytotoxic chemotherapy and blood count recovery in patients.

Although hematopoietic stem cells are generally thought to have the capacity to form all cells of the blood, it is becoming clear that individual stem cells may not be equal in their differentiation potential. That is, some stem cells are “biased” to become mature cells of a particular type. In addition, the general concept of cells having a binary choice of lymphoid or myeloid differentiation is not entirely accurate. A cell population with limited megakaryocytic and erythroid or myeloid (monocyte and granulocyte) and lymphoid potential is now added to the commitment steps stem cells may undergo.

■ SELF-RENEWAL

The hematopoietic stem cell must balance its three potential fates: apoptosis, self-renewal, and differentiation. The proliferation of cells is generally not associated with the ability to undergo a self-renewing division except among memory T and B cells and among stem cells. Self-renewal capacity has generally been regarded as giving way to differentiation as the only option after cell division when cells leave the stem cell compartment, unless they become memory lymphocytes. However, emerging data suggest that some myeloid committed progenitors may also have self-renewing potential *in vivo*, providing long-term production of cells. Stem cells all have self-renewing capacity by definition, and they have an additional feature characterizing their proliferation machinery. Stem cells in many mature adult tissues are heterogeneous with some being deeply quiescent, serving as a deep reserve, whereas others are more proliferative and replenish the short-lived progenitor population. In the hematopoietic system, stem cells are generally cytokine-resistant, remaining dormant even when cytokines drive bone marrow progenitors to proliferation rates measured in hours. Stem cells, in contrast, are thought to divide at far longer intervals, measured in months to years, for the most quiescent cells. This quiescence is difficult to overcome *in vitro*, limiting the ability to effectively expand human hematopoietic stem cells. The process may be controlled by particularly high levels of cyclin-dependent kinase inhibitors like p57 or CDKN1c that restrict entry of stem cells into the cell cycle, blocking the G₁-S transition. Exogenous signals from the niche also appear to enforce quiescence, including angiogenin, interleukin 18, and perhaps angiopoietin 1.

The regulation of stem cell proliferation also appears to change with age. Both cell intrinsic features like the cyclin-dependent kinase inhibitor p16INK4a and bone marrow microenvironment features like declining sympathetic innervation are implicated in age-related stem cell changes. Either lowering expression of p16INK4a or stimulating beta-3 adrenergic receptors in older animals improves stem cell cycling and capacity to reconstitute hematopoiesis in adoptive hosts, making them similar to younger animals. Mature cell numbers are unaffected. Therefore, molecular events governing the specific functions of stem cells are being gradually made clear and offer the potential of new

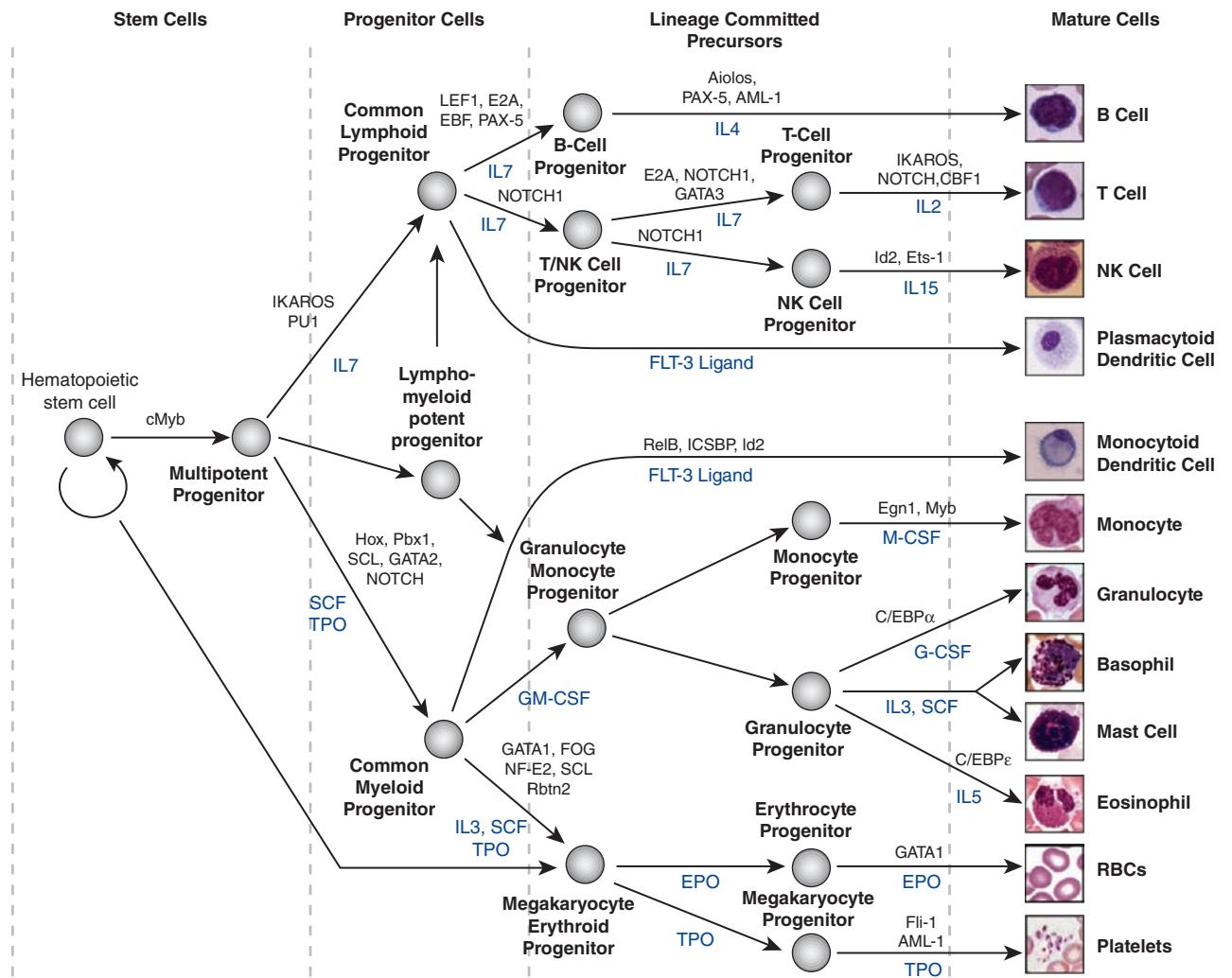


FIGURE 96-2 Hierarchy of hematopoietic differentiation. Stem cells are multipotent cells that are the source of all descendant cells and have the capacity to provide either long-term (measured in years) or short-term (measured in months) cell production. Progenitor cells have a more limited spectrum of cells they can produce and are generally a shorter-lived, highly proliferative population also known as transient amplifying cells. Precursor cells are cells committed to a single blood cell lineage but with a continued ability to proliferate; they do not have all the features of a fully mature cell. Mature cells are the terminally differentiated product of the differentiation process and are the effector cells of specific activities of the blood and immune system. Progress through the pathways is mediated by alterations in gene expression. The regulation of the differentiation by soluble factors and cell-cell communications within the bone marrow niche are still being defined. The transcription factors that characterize particular cell transitions are illustrated on the arrows; the soluble factors that contribute to the differentiation process are in blue. This picture is a simplification of the process. Active research is revealing multiple discrete cell types in the maturation of B cells and T cells and has identified cells that are biased toward one lineage or another (rather than uncommitted) in their differentiation. EPO, erythropoietin; RBC, red blood cell; SCF, stem cell factor; TPO, thrombopoietin.

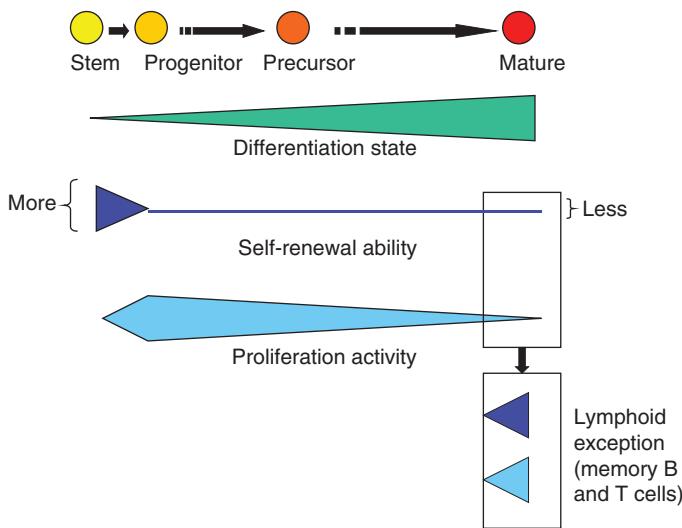


FIGURE 96-3 Relative function of cells in the hematopoietic hierarchy. The boxes represent distinct functional features of cells in the myeloid (upper box) versus lymphoid (lower box) lineages.

therapeutic approaches to changing stem cell functions. One critical stem cell function that remains poorly defined is the molecular regulation of self-renewal.

For medicine, self-renewal is perhaps the most important function of stem cells because it is critical in regulating the number of stem cells. Stem cell number is a key limiting parameter for both autologous and allogeneic stem cell transplantation. Were we to have the ability to use fewer stem cells or expand limited numbers of stem cells ex vivo, it might be possible to reduce the morbidity and expense of stem cell harvests, enable use of other stem cell sources, and improve the potential for gene-modified stem cell transplant. For example, umbilical cord blood is a rich source of stem cells but generally has an inadequate number of stem cells for use in transplantation in adults. These cells have two advantages over other stem cell sources: there is a lower incidence of graft-versus-host disease, and cord blood banks have representation of populations underrepresented in adult donor registries. Hematopoietic reconstitution from cord blood is slow, however, in part due to cell number. Expansion might improve this; however, advances in haploidentical donor cell transplantation have reduced cord blood use.

Gene-modified stem cells are increasingly being tested and have been found to offer great promise for genetic blood diseases like congenital immunodeficiencies and hemoglobinopathies such as sickle cell

disease. However, the complexity and cost of modifying enough stem cells for transplantation is problematic. Expanding a small number of gene-modified stem cells may mitigate that issue. Therefore, understanding self-renewal offers the potential to facilitate development of an important new area of stem cell-based medicine. Some limited understanding of self-renewal exists and, intriguingly, implicates gene products that are associated with the chromatin state, a high-order organization of chromosomal DNA that influences transcription. These include members of the polycomb family, a group of zinc finger-containing transcriptional regulators that interact with the chromatin structure, contributing to the accessibility of groups of genes for transcription. One member, Bmi-1, is important in enabling hematopoietic stem cell self-renewal through modification of cell cycle regulators such as the cyclin-dependent kinase inhibitors. In the absence of Bmi-1 or of the transcriptional regulator, Gfi-1, hematopoietic stem cells decline in number and function. In contrast, dysregulation of Bmi-1 has been associated with leukemia; it may promote leukemic stem cell self-renewal when it is overexpressed. The same is true for the polycomb gene, *Asxl1*, that is commonly mutated in myelodysplasia and leukemia. Other transcription regulators have also been associated with self-renewal, particularly homeobox, or "hox," genes. These transcription factors are named for their ability to govern large numbers of genes, including those determining body patterning in invertebrates. HoxB4 is capable of inducing extensive self-renewal of stem cells through its DNA-binding motif. Other members of the hox family of genes have been noted to affect normal stem cells, but they are also associated with leukemia. Epigenetic modifiers such as the DNA methyl transferase DNMT3a or the dioxygenase involved in DNA demethylation, Tet2, also play a role in stem cell regulation. Like *Asxl1*, mutations of these genes are associated with clonal outgrowth of stem cells bearing the mutations. These mutations are not sufficient for malignancy, but they enable clones bearing them to gain dominance and predispose cells to malignant transformation. They are often referred to as "founder mutations" because myelodysplastic and leukemic cells appear to evolve from them by DNA sequencing analysis.

CANCER IS SIMILAR TO AN ORGAN WITH SELF-RENEWING CAPACITY

The relationship of stem cells to cancer is an important dimension of adult stem cell biology. Cancer may share principles of organization with normal tissues. Cancer cells are heterogeneous even within a given patient and may have a hierarchical organization of cells with a base of stem-like cells capable of the signature stem cell features: self-renewal and differentiation. These stem-like cells might be the basis for perpetuation of the tumor and represent a slowly dividing, rare population with distinct regulatory mechanisms, including a relationship with a specialized microenvironment. A subpopulation of self-renewing cells has been defined for some, but not all, cancers. These include myeloid leukemias where founder mutations appear to enable clones of cells to expand. With additional mutations, these can serve as the initiating or stem cells of a cancer, and eliminating them may be necessary for curing the patient. Understanding the hierarchical cell organization within cancers and whether eliminating cancer stem cell equivalents can improve cure rates is an area of active investigation.

Does the concept of cancer stem cells provide insight into the cellular origin of cancer? The fact that some cells within a cancer have stem cell-like properties does not necessarily mean that the cancer arose in the stem cell itself. Rather, more mature cells could have acquired the self-renewal characteristics of stem cells. Any single genetic event is unlikely to be sufficient to enable full transformation of a normal cell to a frankly malignant one. Rather, cancer is a multistep process, and for the multiple steps to accumulate, the cell of origin must be able to persist for prolonged periods. It must also be able to generate large numbers of daughter cells. The normal stem cell has these properties and, by virtue of its having intrinsic self-renewal capability, may be more readily converted to a malignant phenotype. This hypothesis has been tested experimentally in the hematopoietic system. Taking advantage of the cell-surface markers that distinguish hematopoietic cells of varying maturity, stem cells, progenitors, precursors, and mature cells

can be isolated. Powerful transforming gene constructs were placed in these cells, and it was found that the cell with the greatest potential to produce a malignancy was dependent on the transforming gene. In some cases, it was the stem cell, but in others, the progenitor cell functioned to initiate and perpetuate the cancer. This shows that cells can acquire stem cell-like properties in malignancy.

WHAT ELSE CAN HEMATOPOIETIC STEM CELLS DO?

Some experimental data have suggested that hematopoietic stem cells or other bone marrow cells are capable of playing a role in healing the vascular and tissue damage associated with stroke and myocardial infarction. These data are controversial, and the applicability of a stem cell approach to nonhematopoietic conditions remains experimental. However, reprogramming technology offers the potential for using readily obtained hematopoietic cells as a source for cells with other capabilities. Active areas of investigation are to use reprogrammed cells to generate mature lymphoid cells for immuno-oncology applications or red cells and platelets to overcome dependency on blood donors.

STEM CELLS AS TARGETS OF GENE THERAPY

Tools to alter gene sequence, expression, and regulation are becoming increasingly feasible. The hematopoietic stem cell is a target for a wide range of interventions. Lentiviral, retroviral, and adenoviral vectors are being used to replace defective genes (e.g., in primary immunodeficiency diseases). Antisense technology is being applied to block gene expression (e.g., blocking the *Bcl11a* repression of fetal globin expression in sickle cell disease and thalassemia). CRISPR/Cas technology is being applied to repair abnormal gene sequences. Precision genetic manipulations are expanding, and the hematopoietic system is central to it.

In sum, the stem cell has tremendous healing capacity and is essential for life. However, if dysregulated, it can threaten the life it maintains. Understanding how stem cells function, the signals that modify their behavior, and the tissue niches that modulate stem cell responses to injury and disease is critical for more effectively developing stem cell-based medicines.

FURTHER READING

- ADELMAN ER, FIGUEROA ME: Human hematopoiesis: Aging and leukemogenic risk. *Curr Opin Hematol* 1:57, 2021.
- BARYAWNO N et al: A cellular taxonomy of the bone marrow stroma in homeostasis and leukemia. *Cell* 7:1915, 2019.
- Ito Y et al: Turbulence activates platelet biogenesis to enable clinical scale ex vivo production. *Cell* 3:636, 2018.
- RODRIGUEZ-FRATICELLI AE, CAMARGO F: Systems analysis of hematopoiesis using single-cell lineage tracing. *Curr Opin Hematol* 1:18, 2021.

97

Iron Deficiency and Other Hypoproliferative Anemias

John W. Adamson

Anemias associated with normocytic and normochromic red cells and an inappropriately low reticulocyte response (reticulocyte index <2–2.5) are *hypoproliferative anemias*. This category includes early iron deficiency (before hypochromic microcytic red cells develop), acute and chronic inflammation (including many malignancies), renal disease, hypometabolic states such as protein malnutrition and endocrine

TABLE 97-1 Body Iron Distribution

	IRON CONTENT, mg	
	ADULT MALE, 80 kg	ADULT FEMALE, 60 kg
Hemoglobin	2500	1700
Myoglobin/enzymes	500	300
Transferrin iron	3	3
Iron stores	600–1000	0–300

deficiencies, and anemias from marrow damage. Marrow damage states are discussed in **Chap. 102**.

Hypoproliferative anemias are the most common anemias, and in the clinic, iron deficiency anemia is the most common of these followed by the anemia of inflammation. The anemia of inflammation, similar to iron deficiency, is related in part to abnormal iron metabolism. The anemias associated with renal disease, inflammation, cancer, and hypometabolic states are characterized by a suboptimal erythropoietin response to the anemia.

IRON METABOLISM

Iron is a critical element in the function of all cells, although the amount of iron required by individual tissues varies during development. At the same time, the body must protect itself from free iron, which is highly toxic in that it participates in chemical reactions that generate free radicals such as singlet O_2 or OH^- . Consequently, elaborate mechanisms have evolved that allow iron to be made available for physiologic functions while at the same time conserving this element and handling it in such a way that toxicity is avoided.

The major role of iron in mammals is to carry O_2 as part of hemoglobin. O_2 is also bound by myoglobin in muscle. Iron is a critical element in iron-containing enzymes, including the cytochrome system in mitochondria. Iron distribution in the body is shown in **Table 97-1**. Without iron, cells lose their capacity for electron transport and energy metabolism. In erythroid cells, hemoglobin synthesis is impaired, resulting in anemia and reduced O_2 delivery to tissue.

THE IRON CYCLE IN HUMANS

Figure 97-1 outlines the major pathways of internal iron exchange in humans. Iron absorbed from the diet or released from stores circulates

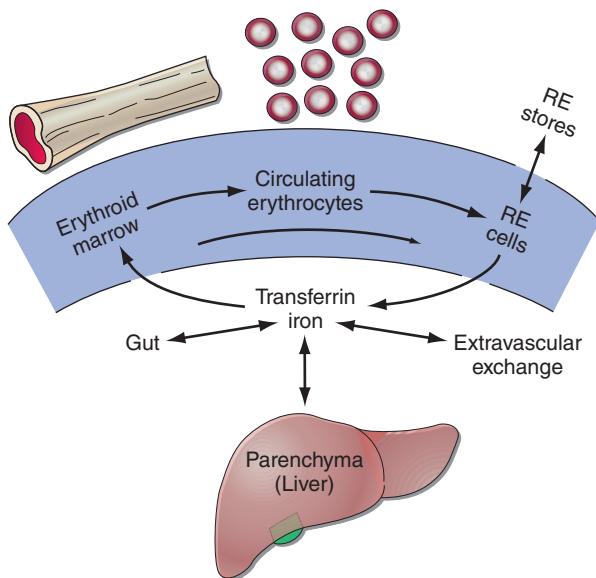


FIGURE 97-1 Internal iron exchange. Normally 80% of iron passing through the plasma transferrin pool is recycled from senescent red cells. Absorption of ~1 mg/d is required from the diet in men, and 1.4 mg/d in women to maintain homeostasis. As long as transferrin saturation is maintained between 20 and 60% and erythropoiesis is not increased, use of iron stores is not required. However, in the event of blood loss, dietary iron deficiency, or inadequate iron absorption, up to 40 mg/d of iron can be mobilized from stores. RE, reticuloendothelial.

in the plasma bound to *transferrin*, the iron transport protein. Transferrin is a bilobed glycoprotein with two iron-binding sites. Transferrin that carries iron exists in two forms—*monoferric* (one iron atom) or *diferric* (two iron atoms). The turnover (half-clearance time) of transferrin-bound iron is very rapid—typically 60–90 min. Because almost all of the iron transported by transferrin is delivered to the erythroid marrow, the clearance time of transferrin-bound iron from the circulation is affected most by the plasma iron level and the erythroid marrow activity. When erythropoiesis is markedly stimulated, the pool of erythroid cells requiring iron increases, and the clearance time of iron from the circulation decreases. The half-clearance time of iron in the presence of iron deficiency is as short as 10–15 min. With suppression of erythropoiesis, the plasma iron level typically increases, and the half-clearance time may be prolonged to several hours. Normally, the iron bound to transferrin turns over 6–8 times per day. Assuming a normal plasma iron level of 80–100 µg/dL, the amount of iron passing through the transferrin pool is 20–24 mg/day.

The iron-transferrin complex circulates in the plasma until it interacts with specific *transferrin receptors* on the surface of marrow erythroid cells. Diferric transferrin has the highest affinity for transferrin receptors; apotransferrin (not carrying iron) has very little affinity. Although transferrin receptors are found on cells in many tissues within the body—and all cells at some time during development will display transferrin receptors—the cell having the greatest number of receptors (300,000–400,000/cell) is the developing erythroblast.

Once the iron-bearing transferrin interacts with its receptor, the complex is internalized via clathrin-coated pits and transported to an acidic endosome, where the iron is released at the low pH. The iron is then made available for heme synthesis while the transferrin-receptor complex is recycled to the surface of the cell, where the bulk of the transferrin is released back into circulation and the transferrin receptor re-anchors into the cell membrane. At this point, a certain amount of the transferrin receptor protein may be released into circulation and can be measured as soluble transferrin receptor protein. Within the erythroid cell, iron in excess of the amount needed for hemoglobin synthesis binds to a storage protein, *apo ferritin*, forming *ferritin*. This mechanism of iron exchange also takes place in other cells of the body expressing transferrin receptors, especially liver parenchymal cells where the iron can be incorporated into heme-containing enzymes or stored. The iron incorporated into hemoglobin subsequently enters the circulation as new red cells are released from the bone marrow. The iron is then part of the red cell mass and will not become available for reutilization until the red cell dies.

In a normal individual, the average red cell life span is 120 days. Thus, 0.8–1% of red cells are replaced each day. At the end of its life span, the red cell is recognized as senescent by the cells of the *reticuloendothelial (RE)* system, and the red cell undergoes phagocytosis. Once within the RE cell, the ingested hemoglobin is broken down, the globin and other proteins are returned to the amino acid pool, and the iron is shuttled back to the surface of the RE cell, where it is presented to circulating transferrin via the iron export channel, ferroportin. It is the efficient and highly conserved recycling of iron from senescent red cells that supports steady-state (and even accelerated) erythropoiesis.

Because each milliliter of red cells contains 1 mg of elemental iron, the amount of iron needed to replace those red cells lost through senescence amounts to 20 mg/d (assuming an adult with a red cell mass of 2 L). Any additional iron required for daily red cell production comes from the diet. Normally, an adult male will need to absorb at least 1 mg of elemental iron daily to meet needs, while females in the childbearing years will need to absorb an average of 1.4 mg/d. However, to achieve a maximum proliferative erythroid marrow response to anemia, additional iron must be available. With markedly stimulated erythropoiesis, demands for iron are increased by as much as six- to eightfold. With extravascular hemolytic anemia, the rate of red cell destruction is increased, but the iron recovered from the red cells is efficiently reutilized for hemoglobin synthesis. In contrast, with intravascular hemolysis or blood loss anemia, the rate of red cell production is limited by the amount of iron that can be mobilized from stores. Typically, the rate of mobilization under these circumstances will not

support red cell production more than 2.5 times normal. If the delivery of iron to the stimulated marrow is suboptimal, the marrow's proliferative response is blunted, and hemoglobin synthesis is impaired. The result is a hypoproliferative marrow accompanied by microcytic, hypochromic anemia.

Whereas blood loss or hemolysis places a demand on the iron supply, inflammatory conditions interfere with iron release from stores and can result in a rapid decrease in the serum iron (see below).

NUTRITIONAL IRON BALANCE

The balance of iron in humans is tightly controlled and designed to conserve iron for reutilization. There is no regulated excretory pathway for iron, and the only mechanisms by which iron is lost are blood loss (via gastrointestinal bleeding, menses, or other forms of bleeding) and the loss of epithelial cells from the skin, gut, and genitourinary tract. Normally, the only route by which iron comes into the body is via absorption from food or from medicinal iron taken orally. Iron may also enter the body through red cell transfusions or injection of iron complexes. The margin between the amount of iron available for absorption and the requirement for iron in growing infants and the adult female is narrow; this accounts for the great prevalence of iron deficiency worldwide—currently estimated at more than one billion people.

The amount of iron required from the diet to replace losses averages ~10% of body iron content a year in men and 15% in women of childbearing age. Dietary iron content is closely related to total caloric intake (~6 mg of elemental iron per 1000 calories). Iron bioavailability is affected by the nature of the foodstuff, with heme iron (e.g., red meat) being most readily absorbed. In the United States, the average iron intake in an adult male is 15 mg/d with 6% absorption; for the average female, the intake is 11 mg/d with 12% absorption. An individual with iron deficiency can increase iron absorption to ~20% of the iron present in a meat-containing diet but only 5–10% of the iron in a vegetarian diet. As a result, one-third of the female population in the United States has virtually no iron stores. Vegetarians are at an additional disadvantage because certain foodstuffs that include phytates and phosphates reduce iron absorption by ~50%. When ionizable iron salts are given together with food, the amount of iron absorbed is reduced. When the percentage of iron absorbed from individual food items is compared with the percentage for an equivalent amount of ferrous salt, iron in vegetables is only about one-twentieth as available, egg iron one-eighth, liver iron one-half, and heme iron one-half to two-thirds.

Infants, children, and adolescents may be unable to maintain normal iron balance because of the demands of body growth and lower dietary intake of iron. During the last two trimesters of pregnancy, daily iron requirements increase to 5–6 mg, and iron supplements are strongly recommended for pregnant women in developed countries.

Iron absorption takes place largely in the duodenum and proximal small intestine and is a carefully regulated process. For absorption, iron must be taken up by the luminal cell. That process is facilitated by the acidic contents of the stomach, which maintains the iron in solution. At the brush border of the absorptive cell, the ferric iron is converted to the ferrous form by a ferrireductase. Transport across the membrane is accomplished by divalent metal transporter type 1 (DMT-1, also known as natural resistance macrophage-associated protein type 2 [Nramp 2] or DCT-1). DMT-1 is a general cation transporter. Once inside the gut cell, iron may be stored as ferritin or transported through the cell to be released at the basolateral surface to plasma transferrin through the membrane-embedded iron exporter, ferroportin. The function of ferroportin is negatively regulated by hepcidin, the principal iron regulatory hormone. In the process of release, iron interacts with another ferroxidase, hephaestin, which oxidizes the iron to the ferric form for transferrin binding. Hephaestin is similar to ceruloplasmin, the copper-carrying protein.

Iron absorption is influenced by a number of physiologic states. Erythroid hyperplasia stimulates iron absorption even in the face of normal or increased iron stores, and hepcidin levels are inappropriately low. Thus, patients with anemias associated with high levels of ineffective erythropoiesis absorb excess amounts of dietary iron. The molecular mechanism underlying this is the production of erythropherrone

(ERFE) by developing erythroblasts. ERFE suppresses hepcidin production, and over time, this may lead to iron overload and tissue damage. In iron deficiency, hepcidin levels are also low and iron is much more efficiently absorbed; the contrary is true in states of secondary iron overload. The normal individual can reduce iron absorption in situations of excessive intake or medicinal iron intake; however, while the percentage of iron absorbed goes down, the absolute amount goes up. This accounts for the acute iron toxicity occasionally seen when children ingest large numbers of iron tablets. Under these circumstances, the amount of iron absorbed exceeds the transferrin binding capacity of the plasma, resulting in free iron that affects critical organs such as cardiac muscle cells.

IRON-DEFICIENCY ANEMIA

Iron deficiency is one of the most prevalent forms of malnutrition. Globally, 50% of anemia is attributable to iron deficiency and accounts for approximately nearly a million deaths annually worldwide. Africa and parts of Asia bear 71% of the global mortality burden; North America represents only 1.4% of the total morbidity and mortality associated with iron deficiency.

STAGES OF IRON DEFICIENCY

The progression to iron deficiency can be divided into three stages (Fig. 97-2). The first stage is *negative iron balance*, in which the demands for (or losses of) iron exceed the body's ability to absorb iron from the diet. This stage results from a number of physiologic mechanisms, including blood loss, pregnancy (in which the demands for red cell production by the fetus outstrip the mother's ability to provide iron), rapid growth spurts in the adolescent, or inadequate dietary iron intake. Blood loss in excess of 10–20 mL of red cells per day is greater than the amount of iron that the gut can absorb from a normal diet. Under these circumstances, the iron deficit must be made up by mobilization of iron from RE storage sites. During this period, iron stores—reflected by the serum ferritin level or the appearance of stainable iron on bone marrow aspirations—decrease. As long as iron stores are present and can be mobilized, the serum iron, total iron-binding

	Normal	Negative iron balance	Iron-deficient erythropoiesis	Iron-deficiency anemia
Iron stores				
Erythron iron				
Marrow iron stores	1-3+	0-1+	0	0
Serum ferritin ($\mu\text{g/L}$)	50-200	<20	<15	<15
TIBC ($\mu\text{g/dL}$)	300-360	>360	>380	>400
SI ($\mu\text{g/dL}$)	50-150	NL	<50	<30
Saturation (%)	30-50	NL	<20	<10
Marrow sideroblasts (%)	40-60	NL	<10	<10
RBC protoporphyrin ($\mu\text{g/dL}$)	30-50	NL	>100	>200
RBC morphology	NL	NL	NL	Microcytic/hypochromic

FIGURE 97-2 Laboratory studies in the evolution of iron deficiency. Measurements of marrow iron stores, serum ferritin, and total iron-binding capacity (TIBC) are sensitive to early iron-store depletion. Iron-deficient erythropoiesis is recognized from additional abnormalities in the serum iron (SI), percent transferrin saturation, the pattern of marrow sideroblasts, and the red blood cell (RBC) protoporphyrin level. Patients with iron-deficiency anemia demonstrate all the same abnormalities plus hypochromic microcytic anemia. (Based on RS Hillman, CA Finch: *The Red Cell Manual*, 7th ed. Philadelphia, F.A. Davis and Co, 1996.)

capacity (TIBC), and red cell protoporphyrin levels remain within normal limits. At this stage, red cell morphology and indices are normal.

When iron stores become depleted, the serum iron begins to fall. Gradually, the TIBC increases, as do red cell protoporphyrin levels. By definition, marrow iron stores are absent when the serum ferritin level is $<15 \mu\text{g/L}$. As long as the serum iron remains within the normal range, hemoglobin synthesis is unaffected despite the dwindling iron stores. Once the transferrin saturation falls to 15–20%, hemoglobin synthesis becomes impaired. This is a period of *iron-deficient erythropoiesis*. Careful evaluation of the peripheral blood smear reveals the first appearance of microcytic cells, and if the laboratory technology is available, one finds hypochromic reticulocytes in circulation. Gradually, the hemoglobin begins to fall, reflecting *iron-deficiency anemia*. The transferrin saturation at this point is $<10\text{--}15\%$.

When moderate anemia is present (hemoglobin 10–13 g/dL), the bone marrow remains hypoproliferative. With more severe anemia (hemoglobin 7–8 g/dL), hypochromia and microcytosis become more prominent, target cells and misshapen red cells (poikilocytes) appear on the blood smear as cigar- or pencil-shaped forms, and the erythroid marrow becomes increasingly ineffective. Consequently, with severe prolonged iron-deficiency anemia, erythroid hyperplasia of the marrow develops, rather than hypoproliferation.

■ CAUSES OF IRON DEFICIENCY

Conditions that increase demand for iron, increase iron loss, or decrease iron intake or absorption can produce iron deficiency (**Table 97-2**).

■ CLINICAL PRESENTATION OF IRON DEFICIENCY

Certain clinical conditions carry an increased likelihood of iron deficiency. Pregnancy, adolescence, periods of rapid growth, and an intermittent history of blood loss of any kind should alert the clinician to possible iron deficiency. A cardinal rule is that the appearance of iron deficiency in an adult male or postmenopausal female means gastrointestinal blood loss until proven otherwise. Signs related to iron deficiency depend on the severity and chronicity of the anemia in addition to the usual signs of anemia—fatigue, pallor, and reduced exercise capacity. *Cheilosis* (fissures at the corners of the mouth) and *koilonychia* (spooning of the fingernails) are signs of advanced tissue iron deficiency. The diagnosis of iron deficiency is typically based on laboratory results.

■ LABORATORY IRON STUDIES

Serum Iron and Total Iron-Binding Capacity The serum iron level represents the amount of circulating iron bound to transferrin. The TIBC is an indirect measure of the circulating transferrin. The normal range for the serum iron is 50–150 µg/dL; the normal range for TIBC is 300–360 µg/dL. Transferrin saturation, which is normally 25–50%, is obtained by the following formula: serum iron $\times 100 \div$

TABLE 97-2 Causes of Iron Deficiency

Increased Demand for Iron

- Rapid growth in infancy or adolescence
- Pregnancy
- Erythropoietin therapy

Increased Iron Loss

- Chronic blood loss
- Menses
- Acute blood loss
- Blood donation
- Phlebotomy as treatment for polycythemia vera

Decreased Iron Intake or Absorption

- Inadequate diet
- Malabsorption from disease (sprue, Crohn's disease)
- Malabsorption from surgery (gastrectomy and some forms of bariatric surgery)
- Acute or chronic inflammation

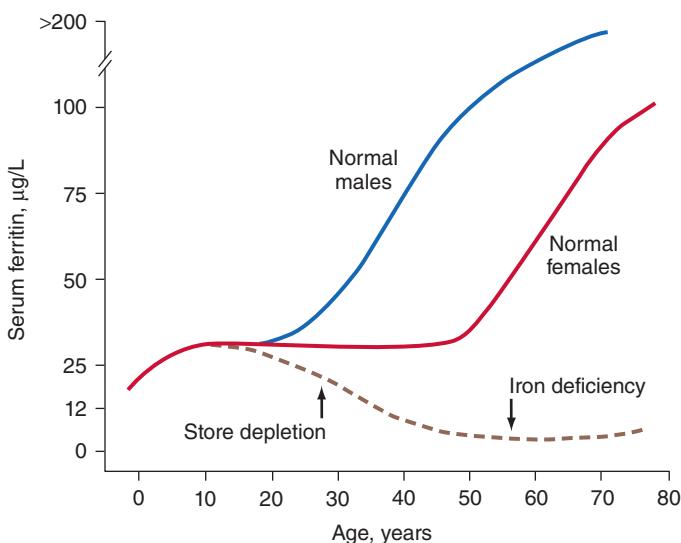


FIGURE 97-3 Serum ferritin levels as a function of sex and age. Iron store depletion and iron deficiency are accompanied by a decrease in serum ferritin level below $20 \mu\text{g/L}$. (Reproduced with permission from RS Hillman: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2011.)

TIBC. Iron-deficiency states are associated with saturation levels $<20\%$. There is a diurnal variation in the serum iron. A transferrin saturation $>50\%$ indicates that a disproportionate amount of the iron bound to transferrin is being delivered to nonerythroid tissues. If this persists for an extended time, tissue iron overload may occur.

Serum Ferritin Free iron is toxic to cells, and the body has established an elaborate set of protective mechanisms to bind iron in various tissue compartments. Within cells, iron is stored complexed to protein as ferritin or hemosiderin. Apoferritin binds to free ferrous iron and stores it in the ferric state. As ferritin accumulates within cells of the RE system, protein aggregates are formed as hemosiderin. Iron in ferritin or hemosiderin can be extracted for release by the RE cells, although hemosiderin is less readily available. Under steady-state conditions, the serum ferritin level correlates with total body iron stores; thus, the serum ferritin level is the most convenient laboratory test to estimate iron stores. The normal value for ferritin varies according to the age and gender of the individual (**Fig. 97-3**). Adult males have serum ferritin values averaging $100 \mu\text{g/L}$, while adult females have levels averaging $30 \mu\text{g/L}$. As iron stores are depleted, the serum ferritin falls to $<15 \mu\text{g/L}$. Such levels are diagnostic of absent body iron stores.

Evaluation of Bone Marrow Iron Stores Although RE iron stores can be estimated from the iron stain of a bone marrow aspirate or biopsy, the measurement of serum ferritin has largely supplanted these procedures for determination of storage iron (**Table 97-3**). The serum ferritin level is a better indicator of iron overload than the marrow iron stain. However, in addition to storage iron, the marrow iron stain provides information about the effective delivery of iron to developing erythroblasts. Normally, when the marrow smear is stained for iron, 20–40% of developing erythroblasts—called *sideroblasts*—will have visible ferritin granules in their cytoplasm. This represents iron in excess of that needed for hemoglobin synthesis. In states in which

TABLE 97-3 Iron Store Measurements

IRON STORES	MARROW IRON STAIN, 0–4+	SERUM FERRITIN, µg/L
0	0	<15
1–300 mg	Trace to 1+	15–30
300–800 mg	2+	30–60
800–1000 mg	3+	60–150
1–2 g	4+	>150
Iron overload	—	$>500\text{--}1000$

release of iron from storage sites is blocked, RE iron will be detectable, and there will be few or no sideroblasts. In the myelodysplastic syndromes, mitochondrial dysfunction can occur, and accumulation of iron in mitochondria appears in a necklace fashion around the nucleus of the erythroblast. Such cells are referred to as *ring sideroblasts*.

Red Cell Protoporphyrin Levels Protoporphyrin is an intermediate in the pathway to heme synthesis. Under conditions in which heme synthesis is impaired, protoporphyrin accumulates within the red cell. This reflects an inadequate iron supply to erythroid precursors to support hemoglobin synthesis. Normal values are <30 µg/dL of red cells. In iron deficiency, values >100 µg/dL are seen. The most common causes of increased red cell protoporphyrin levels are absolute or relative iron deficiency and lead poisoning.

Serum Levels of Transferrin Receptor Protein Because erythroid cells have the highest numbers of transferrin receptors of any cell in the body, and because transferrin receptor protein (TRP) is released by cells into the circulation, serum levels of TRP reflect the total erythroid marrow mass. Another condition in which TRP levels are elevated is absolute iron deficiency. Normal values are 4–9 µg/L determined by immunoassay. This laboratory test is becoming increasingly available and, along with the serum ferritin, has been proposed to distinguish between iron deficiency and the anemia of inflammation (see below).

DIFFERENTIAL DIAGNOSIS

Other than iron deficiency, only three conditions need to be considered in the differential diagnosis of a hypochromic microcytic anemia (**Table 97-4**). The first is an inherited defect in globin chain synthesis: the thalassemias. These are differentiated from iron deficiency most readily by serum iron values; normal or increased serum iron levels and transferrin saturation are characteristic of the thalassemias. In addition, the red blood cell distribution width (RDW) index is generally normal in thalassemia and elevated in iron deficiency.

The second condition is the anemia of inflammation (AI; also referred to as the anemia of chronic disease) with inadequate iron supply to the erythroid marrow. The distinction between true iron-deficiency anemia and AI is among the most common diagnostic problems encountered by clinicians (see below). Usually, AI is normocytic and normochromic. The iron values usually make the differential diagnosis clear, as the ferritin level is normal or increased and the percent transferrin saturation and TIBC are typically below normal.

Finally, the myelodysplastic syndromes represent the third and least common condition. Occasionally, patients with myelodysplasia have impaired hemoglobin synthesis with mitochondrial dysfunction, resulting in impaired iron incorporation into heme. The iron values again reveal normal stores and more than an adequate supply to the marrow, despite the microcytosis and hypochromia.

TREATMENT

Iron-Deficiency Anemia

The severity and cause of iron-deficiency anemia will determine the appropriate approach to treatment. As an example, symptomatic elderly patients with severe iron-deficiency anemia and

TABLE 97-5 Oral Iron Preparations

GENERIC NAME	TABLET (IRON CONTENT), mg	ELIXIR (IRON CONTENT), mg IN 5 mL
Ferrous sulfate	325 (65)	300 (60)
	195 (39)	90 (18)
Extended release	525 (105)	
Ferrous fumarate	325 (107) 195 (64)	100 (33)
Ferrous gluconate	325 (39)	300 (35)
Polysaccharide iron	150 (150) 50 (50)	100 (100)

cardiovascular instability may require red cell transfusions. Younger individuals who have compensated for their anemia can be treated more conservatively with iron replacement. The foremost issue for the latter patient is the precise identification of the cause of the iron deficiency.

For the majority of cases of iron deficiency (pregnant women, growing children and adolescents, patients with infrequent episodes of bleeding, and those with inadequate dietary intake of iron), oral iron therapy will suffice. For patients with unusual blood loss or malabsorption, specific diagnostic tests and appropriate therapy take priority. Once the diagnosis of iron-deficiency anemia and its cause is made, there are three major therapeutic approaches.

RED CELL TRANSFUSION

Transfusion therapy is reserved for individuals who have symptoms of anemia, cardiovascular instability, and continued and excessive blood loss from whatever source and who require immediate intervention. The management of these patients is less related to the iron deficiency than it is to the consequences of the severe anemia. Not only do transfusions correct the anemia acutely, but the transfused red cells provide a source of iron for reutilization, assuming they are not lost through continued bleeding. Transfusion therapy will stabilize the patient while other options are reviewed.

ORAL IRON THERAPY

In the asymptomatic patient with established iron-deficiency anemia and an intact gastrointestinal tract, treatment with oral iron is usually adequate. Encouraging dietary intake of iron-rich foods is also useful. Such foods include oysters, kidney beans, beef liver, tofu, beef (chuck roast, lean ground beef), turkey leg, whole-wheat bread, tuna, eggs, shrimp, peanut butter, leg of lamb, brown rice, raisin bran (whole grain-enriched cereals), lentils, and beans. Multiple preparations of oral iron supplements are available, ranging from simple iron salts to complex iron compounds designed for sustained release throughout the small intestine (**Table 97-5**). Although the various preparations contain different amounts of iron, they are generally all absorbed well and are effective in treatment. Some come with other compounds designed to enhance iron absorption, such as ascorbic acid. It is not clear whether the benefits of such compounds justify their costs. Typically, for iron replacement therapy, up to 200 mg of elemental iron per day is given, usually

TABLE 97-4 Diagnosis of Microcytic Anemia

TESTS	IRON DEFICIENCY	INFLAMMATION	THALASSEMIA	SIDEROBLASTIC ANEMIA
Smear	Micro/hypo	Normal micro/hypo	Micro/hypo with targeting	Variable
Serum iron (µg/dL)	<30	<50	Normal to high	Normal to high
TIBC (µg/dL)	>360	<300	Normal	Normal
Percent saturation	<10	10–20	30–80	30–80
Ferritin (µg/L)	<15	30–200	50–300	50–300
Hemoglobin pattern on electrophoresis	Normal	Normal	Abnormal with β thalassemia; can be normal with α thalassemia	Normal

Abbreviation: TIBC, total iron-binding capacity.

as three or four iron tablets (each containing 50–65 mg elemental iron) given over the course of the day. Ideally, oral iron preparations should be taken on an empty stomach, since food may inhibit iron absorption. Some patients with gastric disease or prior gastric surgery require special treatment with iron solutions because the retention capacity of the stomach may be reduced. The retention capacity is necessary for dissolving the shell of the iron tablet before the release of iron. A dose of 200 mg of elemental iron per day should result in the absorption of iron up to 50 mg/d. This supports a red cell production level of two to three times normal in an individual with a normally functioning marrow and appropriate erythropoietin (EPO) stimulus. However, as the hemoglobin level rises, EPO stimulation decreases, and the amount of iron absorbed is reduced. The goal of therapy in individuals with iron-deficiency anemia is not only to repair the anemia, but also to provide stores of at least 0.5–1 g of iron. Sustained treatment for a period of 6–12 months after correction of the anemia will be necessary to achieve this.

Of the complications of oral iron therapy, gastrointestinal distress is the most prominent and is seen in at least 15–20% of patients. Abdominal pain, nausea, vomiting, or constipation may lead to noncompliance. Although small doses of iron or iron preparations with delayed release may help somewhat, the gastrointestinal side effects are a major impediment to the effective treatment of a number of patients.

The response to iron therapy varies, depending on the EPO stimulus and the rate of absorption. Typically, the reticulocyte count should begin to increase within 4–7 days after initiation of therapy and peak at 1–1½ weeks. The absence of a response may be due to poor absorption, noncompliance (which is common), or a confounding diagnosis. A useful test in the clinic to determine the patient's ability to absorb iron is the *iron tolerance test*. Two iron tablets are given to the patient on an empty stomach, and the serum iron is measured serially over the subsequent 2–3 h. Normal absorption will result in an increase in the serum iron of at least 100 µg/dL. If iron deficiency persists despite adequate treatment, it may be necessary to switch to parenteral iron therapy.

PARENTERAL IRON THERAPY

Intravenous iron can be given to patients who are unable to tolerate oral iron; whose needs are relatively acute; or who need iron on an ongoing basis, usually due to persistent gastrointestinal or menstrual blood loss. Parenteral iron use has been increasing rapidly over the past several years with the recognition that recombinant EPO therapy induces a large demand for iron—a demand that frequently cannot be met through the physiologic release of iron from RE sources or oral iron absorption. The safety of parenteral iron has been a concern largely driven by the high adverse reaction rate to high-molecular-weight iron dextran. The newer iron complexes that are available, such as ferumoxytol (Feraheme), sodium ferric gluconate (Ferrlecit), iron sucrose (Venofer), low-molecular-weight (LMW) iron dextran (InFed), ferric derisomaltose (Monoferric), and ferric carboxymaltose (Injectafer), have much lower rates of adverse effects. Ferumoxytol delivers 510 mg of iron per infusion; ferric gluconate 125 mg per infusion; LMW iron dextran up to 1500 mg per infusion; ferric carboxymaltose 750 mg per infusion; ferric derisomaltose 1000 mg per infusion; and iron sucrose 200 mg per infusion.

Parenteral iron is used in two ways: one is to administer the total dose of iron required to correct the hemoglobin deficit and provide the patient with at least 500 mg of iron stores; the second is to give repeated small doses of parenteral iron over a protracted period. The latter approach is common in dialysis centers, where it is not unusual for 100 mg of elemental iron to be given weekly for 10 weeks to augment the response to recombinant EPO therapy. The amount of iron needed by an individual patient is calculated by the following formula:

$$\text{Body weight (kg)} \times 2.3 \times (15 - \text{patient's hemoglobin, g/dL}) \\ + 500 \text{ or } 1000 \text{ mg (for stores)}$$

In administering any intravenous iron preparation, anaphylaxis is a concern. Anaphylaxis is much rarer with the newer preparations. The factors that have correlated with an anaphylactic-like reaction include a history of multiple allergies or a prior allergic reaction to an iron preparation. Generalized symptoms appearing several days after the infusion of a large dose of iron can include arthralgias, skin rash, and low-grade fever. These may be dose-related, but they do not preclude the further use of parenteral iron in the patient. To date, patients with sensitivity to one iron preparation have been safely treated with other parenteral iron preparations. If a large dose of LMW iron dextran is to be given (>100 mg), the iron preparation should be diluted in 5% dextrose in water or 0.9% NaCl solution. The iron solution can then be infused over a 60- to 90-min period (for larger doses) or at a rate convenient for the attending nurse or physician. Although a test dose (25 mg) of parenteral LMW iron dextran is recommended, in reality, a slow infusion of a larger dose of parenteral iron solution will afford the same kind of early warning as a separately injected test dose. Early in the infusion of iron, if chest pain, wheezing, a fall in blood pressure, or other systemic symptoms occur, the infusion of iron should be stopped immediately.

OTHER HYPOPROLIFERATIVE ANEMIAS

In addition to mild to moderate iron-deficiency anemia, the hypoproliferative anemias can be divided into four categories: (1) chronic inflammation, (2) renal disease, (3) endocrine and nutritional deficiencies (hypometabolic states), and (4) marrow damage (Chap. 102). With chronic inflammation, renal disease, or hypometabolism, endogenous EPO production is inadequate for the degree of anemia observed. For the anemia of chronic inflammation, the erythroid marrow also responds inadequately to stimulation, due in part to defective *iron reutilization*. As a result of the lack of adequate EPO stimulation, an examination of the peripheral blood smear will disclose only an occasional polychromatophilic ("shift") reticulocyte. In cases of iron deficiency or marrow damage, appropriate elevations in endogenous EPO levels are typically found, and shift reticulocytes will be present on the blood smear.

■ ANEMIA OF ACUTE AND CHRONIC INFLAMMATION/INFECTION (AI)

AI, which encompasses inflammation, infection, tissue injury, and conditions (e.g., cancer) associated with the release of proinflammatory cytokines, is one of the most common forms of anemia seen clinically. It is the most important anemia in the differential diagnosis of iron deficiency because many of the features of the anemia are brought about by inadequate iron delivery to the marrow, despite the presence of normal or increased iron stores. This is reflected by a low serum iron, increased red cell protoporphyrin, a hypoproliferative marrow, transferrin saturation in the range of 15–20%, and a normal or increased serum ferritin. The serum ferritin values are often the most distinguishing features between true iron-deficiency anemia and the iron-restricted erythropoiesis associated with inflammation. Typically, serum ferritin values increase threefold over basal levels in the face of inflammation. These changes are due to the effects of inflammatory cytokines and hepcidin, the key iron regulatory hormone, acting at several levels of erythropoiesis (Fig. 97-4).

Interleukin 1 (IL-1) directly decreases EPO production in response to anemia. IL-1, acting through accessory cell release of interferon γ (IFN- γ), suppresses the response of the erythroid marrow to EPO—an effect that can be overcome by EPO administration in vitro and in vivo. In addition, tumor necrosis factor (TNF), acting through the release of IFN- β by marrow stromal cells, also suppresses the response to EPO. Hepcidin, made by the liver, is increased in inflammation via an IL-6-mediated pathway, and acts to suppress iron absorption and iron release from storage sites. The overall result is a chronic hypoproliferative anemia with classic changes in iron metabolism. The anemia is further compounded by a mild to moderate shortening in red cell survival.

With chronic inflammation, the primary disease will determine the severity and characteristics of the anemia. For example, many

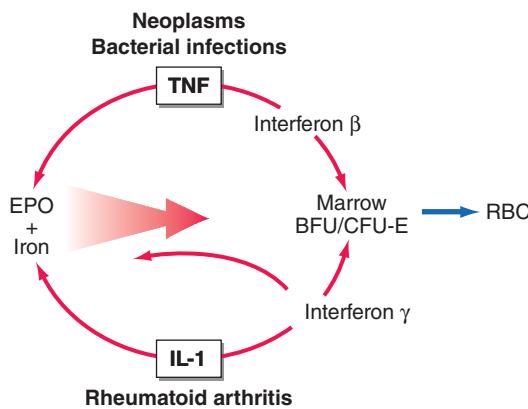


FIGURE 97-4 Suppression of erythropoiesis by inflammatory cytokines. Through the release of tumor necrosis factor (TNF) and interferon β (IFN- β), neoplasms and bacterial infections suppress erythropoietin (EPO) production and the proliferation of erythroid progenitors (erythroid burst-forming units and erythroid colony-forming units [BFU/CFU-E]). The mediators in patients with vasculitis and rheumatoid arthritis include interleukin 1 (IL-1) and IFN- γ . The red arrows indicate sites of inflammatory cytokine inhibitory effects. RBC, red blood cell.

patients with cancer also have anemia that is typically normocytic and normochromic. In contrast, patients with long-standing active rheumatoid arthritis or chronic infections such as tuberculosis will have a microcytic, hypochromic anemia. In both cases, the bone marrow is hypoproliferative, but the differences in red cell indices reflect differences in the availability of iron for hemoglobin synthesis. Occasionally, conditions associated with chronic inflammation are also associated with chronic blood loss. Under these circumstances, the measurement of soluble transferrin receptor protein may be necessary to rule out absolute iron deficiency. However, the administration of iron in this case will correct the iron-deficiency component of the anemia and leave the inflammatory component unaffected.

The anemia associated with acute infection or inflammation is typically mild but becomes more pronounced over time. Acute infection can produce a decrease in hemoglobin levels of 2–3 g/dL within 1 or 2 days; this is largely related to the hemolysis of red cells near the end of their natural life span. The fever and cytokines released exert a selective pressure against cells with more limited capacity to maintain the red cell membrane. In most individuals, the mild anemia is reasonably well tolerated, and symptoms, if present, are associated with the underlying disease. Occasionally, in patients with preexisting cardiac disease, moderate anemia (hemoglobin 10–11 g/dL) may be associated with angina, exercise intolerance, and shortness of breath. The erythropoietic profile that distinguishes the anemia of inflammation from the other causes of hypoproliferative anemias is shown in Table 97-6.

■ ANEMIA OF CHRONIC KIDNEY DISEASE (CKD)

Progressive CKD is usually associated with a moderate to severe hypoproliferative anemia; the level of the anemia correlates with the stage of CKD. Red cells are typically normocytic and normochromic, and reticulocytes are decreased. The anemia is primarily due to a failure of EPO production by the diseased kidney and a reduction in red cell survival. In certain forms of acute renal failure, the correlation

between the anemia and renal function is weaker. Patients with the hemolytic-uremic syndrome increase erythropoiesis in response to the hemolysis, despite renal failure. Polycystic kidney disease also shows a smaller degree of EPO deficiency for a given level of renal failure. By contrast, patients with diabetes or myeloma have more severe EPO deficiency for a given level of renal failure.

Assessment of iron status provides information to distinguish the anemia of CKD from the other forms of hypoproliferative anemia (Table 97-6) and to guide management. Patients with the anemia of CKD usually present with normal serum iron, TIBC, and ferritin levels. However, those maintained on chronic hemodialysis may develop iron deficiency from blood loss through the dialysis procedure. Iron must be replenished in these patients to ensure an adequate response to EPO therapy (see below).

■ ANEMIA IN HYPOMETABOLIC STATES

Patients who are starving, particularly for protein, and those with a variety of endocrine disorders that produce lower metabolic rates, may develop a mild to moderate hypoproliferative anemia. The release of EPO from the kidney is sensitive to the need for O_2 , not just O_2 levels. Thus, EPO production is triggered at lower levels of blood O_2 content in disease states (e.g., hypothyroidism and starvation) where metabolic activity, and thus O_2 demand, is decreased.

Endocrine Deficiency States The difference in the levels of hemoglobin between men and women is related to the effects of androgen and estrogen on erythropoiesis. Testosterone and anabolic steroids augment erythropoiesis; castration and estrogen administration to males decrease erythropoiesis. Patients who are hypothyroid or have deficits in pituitary hormones also may develop a mild anemia. Pathogenesis may be complicated by other nutritional deficiencies because iron and folic acid absorption can be affected by these disorders. Usually, correction of the hormone deficiency reverses the anemia.

Anemia may be more severe in Addison's disease, depending on the level of thyroid and androgen hormone dysfunction; however, anemia may be masked by decreases in plasma volume. Once such patients are given cortisol and volume replacement, the hemoglobin level may fall rapidly. Mild anemia complicating hyperparathyroidism may be due to decreased EPO production as a consequence of the renal effects of hypercalcemia or to impaired proliferation of erythroid progenitors.

Protein Starvation Decreased dietary intake of protein may lead to mild to moderate hypoproliferative anemia; this form of anemia may be prevalent in the elderly. The anemia can be more severe in patients with a greater degree of starvation. In marasmus, where patients are both protein- and calorie-deficient, the release of EPO is impaired in proportion to the reduction in metabolic rate; however, the degree of anemia may be masked by volume depletion and becomes apparent after refeeding. Deficiencies in other nutrients (iron, folate) may also complicate the clinical picture but may not be apparent at diagnosis. Changes in the erythrocyte indices on refeeding should prompt evaluation of iron, folate, and B_{12} status.

Anemia in Liver Disease A mild hypoproliferative anemia may develop in patients with chronic liver disease from nearly any cause. The peripheral blood smear may show spur cells and stomatocytes

TABLE 97-6 Diagnosis of Hypoproliferative Anemias

TESTS	IRON DEFICIENCY	INFLAMMATION	RENAL DISEASE	HYPOMETABOLIC STATES
Anemia	Mild to severe	Mild	Mild to severe	Mild
MCV (fL)	60–90	80–90	90	90
Morphology	Normo-microcytic	Normocytic	Normocytic	Normocytic
SI (μ g/dL)	<30	<50	Normal	Normal
TIBC (μ g/dL)	>360	<300	Normal	Normal
Saturation (%)	<10	10–20	Normal	Normal
Serum ferritin (μ g/L)	<15	30–200	115–150	Normal
Iron stores	0	2–4+	1–4+	Normal

Abbreviations: MCV, mean corpuscular volume; SI, serum iron; TIBC, total iron-binding capacity.

from the accumulation of excess cholesterol in the membrane from a deficiency of lecithin-cholesterol acyltransferase. Red cell survival is shortened, and the production of EPO is inadequate to compensate. In alcoholic liver disease, nutritional deficiencies are common and complicate the management. Folate deficiency from inadequate intake, as well as iron deficiency from blood loss and inadequate intake, can alter the red cell indices.

ANEMIA IN AGING

Anemia is common in people over age 65 years. It has been estimated to affect ~11% of community-living older adults and up to 40% of nursing home residents. In at least one-third of these anemic people, a cause for the anemia is not found. Patients with the unexplained anemia of aging do not have nutrient deficiency or renal dysfunction, and although older people can have an increase in systemic inflammatory cytokines (the inflammation of aging), the levels are not high enough to mimic the anemia of chronic inflammation. If hepcidin levels are elevated at all, they are minimally so.

Investigations into the cause(s) of this form of anemia have noted that EPO levels are generally in the normal range, that is, they are inappropriately low for the hemoglobin level. In general, in older people who maintain a normal hemoglobin level, EPO levels increase with age. This compensatory increase to maintain normal oxygen delivery seems to be due to a relative resistance to EPO stimulation; studies of red cell life span in older people have not noted a decrease in red cell survival. More data on the mechanism are needed.

The importance of this unexplained anemia of aging is that low hemoglobin levels are associated with increases in falls, hospitalizations, development of frailty, and mortality. It is not clear whether reversing the anemia would influence these increased risks. Anecdotal evidence suggests that this form of anemia is responsive to exogenous EPO.

TREATMENT

Hypoproliferative Anemias

Many patients with hypoproliferative anemias experience recovery of normal hemoglobin levels when the underlying disease is appropriately treated. For those in whom such reversals are not possible—such as patients with end-stage kidney disease, cancer, and chronic inflammatory diseases—symptomatic anemia requires treatment. The two major forms of treatment are transfusions and EPO.

TRANSFUSIONS

Thresholds for transfusion should be determined based on the patient's symptoms. In general, patients without serious underlying cardiovascular or pulmonary disease can tolerate hemoglobin levels above 7–8 g/dL and do not require intervention until the hemoglobin falls below that level. Patients with more physiologic compromise may need to have their hemoglobin levels kept above 11 g/dL. Usually, a unit of packed red cells increases the hemoglobin level by 1 g/dL. Transfusions are associated with certain infectious risks (*Chap. 113*), and chronic transfusions can produce iron overload. Importantly, the liberal use of blood has been associated with increased morbidity and mortality, particularly in the intensive care setting. Therefore, in the absence of documented tissue hypoxia, a conservative approach to the use of red cell transfusions is preferable.

ERYTHROPOIETIN

EPO is particularly useful in anemias in which endogenous EPO levels are inappropriately low, such as CKD or AI. Iron status must be evaluated and iron replaced to obtain optimal effects from EPO. In patients with CKD, the usual dose of EPO is 50–150 U/kg three times a week intravenously. Hemoglobin levels of 10–12 g/dL are usually reached within 4–6 weeks if iron levels are adequate; 90% of these patients respond. Once a target hemoglobin level is achieved, the EPO dose can be decreased. A decrease in hemoglobin level occurring in the face of EPO therapy usually signifies the development of an infection or iron depletion. Aluminum toxicity

and hyperparathyroidism can also compromise the response to EPO. When an infection intervenes, it is best to interrupt the EPO therapy and rely on transfusions to correct the anemia until the infection is adequately treated. The dose of EPO needed to correct chemotherapy-induced anemia in patients with cancer is higher, up to 300 U/kg three times a week, and only ~60% of patients respond. Because of evidence that there is an increased risk of thromboembolic complications and tumor progression with EPO administration, the risks and benefits of using EPO in such patients must be weighed carefully, and the target hemoglobin should be that necessary to avoid transfusions.

Longer-acting preparations of EPO can reduce the frequency of injections. Darbepoetin alfa, a molecularly modified EPO with additional carbohydrate, has a half-life in the circulation that is three to four times longer than recombinant human EPO, permitting weekly or every other week dosing.

Orally bioavailable EPO mimetics such as roxadustat (usual dose 50 mg PO thrice weekly) that act to increase the biological half-life of active hypoxia-inducible factor (HIF) are demonstrating activity to increase hemoglobin levels in patients with chronic renal disease and other settings.

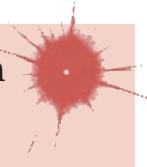
FURTHER READING

- ANDREWS N: Forging a field. The golden age of iron biology. *Blood* 112:219, 2008.
- AUERBACH M, ADAMSON J: How we diagnose and treat iron deficiency anemia. *Am J Hematol* 91:31, 2016.
- DRÜKE T, PARFREY P: Summary of the KDIGO guideline on anemia and comment: Reading between the (guide)line(s). *Kidney Int* 82:952, 2012.
- GANZ T: Anemia of inflammation. *N Engl J Med* 381:1149, 2019.
- KAUTZ L et al: Identification of erythroferrone as an erythroid regulator of iron metabolism. *Nat Genet* 46:678, 2014.
- KRAYENBUEHL P-A et al: Intravenous iron for the treatment of fatigue in non-anemic, premenopausal woman with low serum ferritin concentration. *Blood* 118:3222, 2011.
- PUNNONEN K et al: Serum transferrin receptor and its ratio to serum ferritin in the diagnosis of iron deficiency. *Blood* 89:1052, 1997.

98

Disorders of Hemoglobin

Martin H. Steinberg



Hemoglobinopathies affect the amino acid sequence of globin; thalassemia is a disorder of reduced globin biosynthesis. Together, these disorders of the hemoglobin molecule are our most common Mendelian genetic diseases. They are responsible for most cases of hemolytic anemia. Sickle cell disease and the hemoglobin E (HbE)-associated syndromes are the most prevalent hemoglobinopathies; β and α thalassemia are the most prevalent thalassemias. In addition to these common disorders of hemoglobin, rare globin mutations can cause hemoglobin instability, increased or decreased affinity of hemoglobin for oxygen (O_2), and oxidized hemoglobin reducing O_2 transport. O_2 transport by hemoglobin can also be reduced by exposure to carbon monoxide (CO) and some oxidizing agents (*Table 98-1*).

Phenotypic diversity among hemoglobin disorders is enormous. Mutations can be asymptomatic, for example in heterozygous carriers of sickle hemoglobin (HbS) and thalassemia, or cause intrauterine death as when all α -globin genes are deleted. Impressive gains in understanding the biological basis of hemoglobinopathies and thalassemia have led to novel therapeutics with the promise of improved patient outcomes.

TABLE 98-1 Disorders of Hemoglobin

I. Hemoglobinopathies —hemoglobin variants with amino acid sequence variants that alter the physical, chemical, or functional properties of hemoglobin	Hemoglobin is a tetramer of two pairs of unlike globin polypeptide chains, each chain containing a tetrapyrrole heme group. O ₂ binds to heme as erythrocytes traverse the lungs and is released in the tissues. Heme is nestled within a protective pocket of each globin subunit.
A. Common variants with unusual properties	
1. HbS—polymerization	
2. HbE—reduced biosynthesis	
3. HbC—hemoglobin-membrane interaction	
B. Altered oxygen affinity	
1. High affinity—erythrocytosis	
2. Low affinity—cyanosis, anemia	
C. Hemoglobins that oxidize readily	
1. Unstable hemoglobins—hemolytic anemia, jaundice	
2. M hemoglobins—methemoglobinemia, cyanosis	
II. Thalassemias —defective biosynthesis of globin chains	
A. α Thalassemias	
B. β Thalassemias	
C. Complex thalassemias	
III. Hereditary persistence of fetal hemoglobin —persistence of higher than normal levels of HbF into adult life	
A. Deletions within the <i>HBB</i> cluster—15–30% HbF in heterozygotes, pancellular	
B. Point mutations in <i>HBG2/1</i> promoters—5–30% HbF in heterozygotes; pancellular or heterocellular	
IV. Acquired hemoglobinopathies	
A. Methemoglobin due to toxic exposures	
B. Sulfhemoglobin due to toxic exposures	
C. Carboxyhemoglobin	
D. HbH in erythroleukemia	
E. Elevated HbF in myelodysplasia	

HEMOGLOBIN

Easy access to erythrocytes to study hemoglobin structure and function, reticulocytes to examine hemoglobin biosynthesis, and leukocyte DNA to define the mutations of hemoglobin and the availability of hematopoietic stem and progenitor cells from blood and bone marrow have placed hemoglobin disorders in the forefront of molecular medicine. A review of the biology of hemoglobin provides the background for understanding the pathophysiology of its many genetic and acquired disorders and approaches to their treatment.

■ DEVELOPMENTAL BIOLOGY

Successive waves of erythropoiesis beginning in the yolk sac, moving to the fetal liver and bone marrow, and culminating in the adult marrow direct the synthesis of different hemoglobin molecules that result from sequential activation and silencing of the globin genes (Fig. 98-1).

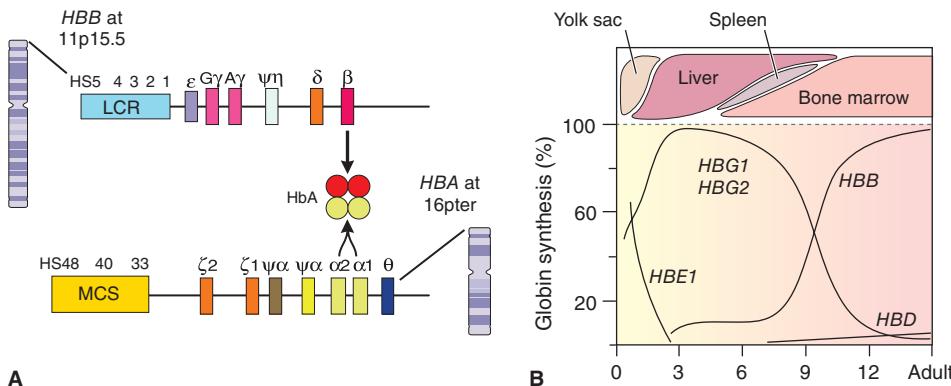


FIGURE 98-1 Globin gene clusters and their hemoglobin products during gestation. **A**, The order of globin genes in the β- and α-globin gene clusters along with their upstream enhancers, the locus control region (LCR) and multispecies conserved sequences (MCS). Normal hemoglobin tetramers contain two α-globin chains and two non-α-globin chains. In the example shown, this is adult HbA. **B**, Sites of erythropoiesis and globin synthesized from the yolk sac and the early embryo (months 1–3), the fetus (months 3–9), after delivery (months 9–12), and afterward (adult).

Hemoglobin is a tetramer of two pairs of unlike globin polypeptide chains, each chain containing a tetrapyrrole heme group. O₂ binds to heme as erythrocytes traverse the lungs and is released in the tissues. Heme is nestled within a protective pocket of each globin subunit.

■ GLOBIN GENE CLUSTERS

Globin is encoded in two nonallelic gene clusters. The β-globin gene cluster is on the short arm of chromosome 11 (11p15.4); the α-globin gene cluster is on chromosome 16 (16p13.3) (Fig. 98-1). The β-globin gene cluster contains an embryonic ε-globin gene (*HBE*), two nearly identical fetal γ-globin genes (*HBG2*, *HBG1*) a major adult β-globin gene (*HBB*), and a minor adult δ-globin gene (*HBD*). The α-globin gene cluster contains an embryonic ζ-globin gene (*HBZ*) and duplicated α-globin genes (*HBA2*, *HBA1*) with identical proteins. Embryonic hemoglobins include Gower I ($\zeta_2\epsilon_2$), Gower II ($\alpha_2\epsilon_2$), Portland I ($\zeta_2\gamma_2$), and Portland II ($\zeta_2\beta_2$). Fetal hemoglobin (HbF, $\alpha_2\gamma_2$) production begins at 6–8 weeks' gestation, peaks during mid-gestation, then falls to <1% of total hemoglobin during the first 6 months of extrauterine life. Adult hemoglobin A (HbA; $\alpha_2\beta_2$) production follows a pattern reciprocal to that of HbF. The hemoglobin composition of normal adults is >95% HbA, ~1% HbF, and 2–3% HbA₂ ($\alpha_2\delta_2$). In adults, HbF and HbA₂ have little functional significance because of their low concentrations, although they can be diagnostically important. Hemoglobin is also subject to posttranslational modifications, the most important being the nonenzymatic glycosylation of HbA forming the adduct HbA_{1c}, which is of diagnostic utility in the management of diabetes mellitus.

■ HEMOGLOBIN STRUCTURE

All globin polypeptides have similar but not identical primary structures. α-Globins contain 141 amino acids, and β-like globins have 146 amino acids. This primary structure dictates, according to the constraints of protein folding, the secondary structure of globin into α-helical sections joined by small nonhelical stretches. Each globin chain folds into a tertiary conformation known as the globin fold, whereby charged amino acid residues face the exterior of the molecules and uncharged residues face the hydrophobic interior. The iron-containing tetrapyrrole heme moiety is protected from oxidation and located between two of the helical segments; O₂ loading and unloading occur when heme iron is in its reduced ferrous form. Globin gene mutations affecting critical heme-binding amino acid residues allow iron to be oxidized, forming methemoglobin, which has high O₂ affinity and does not release O₂ in tissues. Dimers of α- and non-α-globin chains reversibly assemble into tetramers, forming a quaternary structure.

■ HEMOGLOBIN FUNCTION

Hemoglobin transports O₂ from lungs to tissues and carbon dioxide (CO₂) from tissues to lungs and is a nitrate reductase that releases nitric oxide (NO) from nitrite to promote vasodilation. Oxygen binding is defined by the sigmoidal shape of the hemoglobin-O₂ dissociation curve. P₅₀ is a point on this curve that indicates the partial pressure of O₂ where hemoglobin is half saturated (Fig. 98-2). The P₅₀ is influenced by the binding of 2,3-bisphosphoglycerate, a product of glycolysis, in the central cavity of hemoglobin, and by pH and temperature. Normal P₅₀ is ~26 mmHg; low P₅₀ indicates that hemoglobin has high O₂ affinity, decreasing O₂ delivery to tissues; high P₅₀ indicates that hemoglobin has low O₂ affinity, releasing more O₂ to tissues. The conformation of hemoglobin fully saturated with O₂ is known as the R or relaxed state; desaturated hemoglobin is in the T or tense state. The transition between T and R states occurs when two or three O₂ molecules are bound. Cooperativity describes the progressively more rapid binding of O₂ once the first molecule

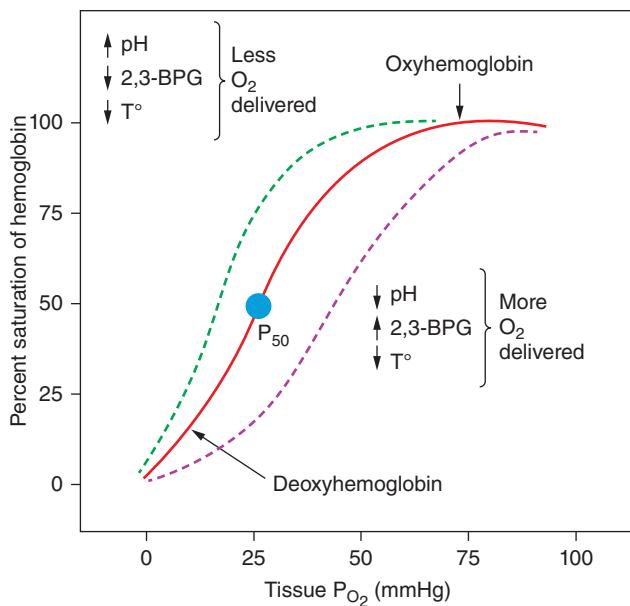


FIGURE 98-2 Hemoglobin-oxygen dissociation curve. The hemoglobin tetramer can bind up to four molecules of oxygen (O_2) in the iron-containing sites of the heme molecules. As O_2 is bound, 2,3-bisphosphoglycerate (2,3-BPG) and carbon dioxide (CO_2) are expelled. Salt bridges are broken, and each of the globin molecules changes its conformation to facilitate O_2 binding. O_2 release to the tissues is the reverse process, with salt bridges being formed and 2,3-BPG and CO_2 bound. Deoxyhemoglobin does not bind O_2 efficiently until the cell returns to conditions of higher pH, the most important modulator of O_2 affinity (Bohr effect). When acid is produced in the tissues, the dissociation curve shifts to the right, facilitating O_2 release and CO_2 binding. Alkalosis has the opposite effect, reducing O_2 delivery.

is bound. Hemoglobin variants that decrease P_{50} are characterized by isolated erythrocytosis as compensation for hypoxia; variants with increased P_{50} sometimes are accompanied by cyanosis and anemia as hemoglobin becomes unsaturated and O_2 delivery is enhanced. Mutations of residues critical for heme binding, R-T transitions, or tetramer stability cause hemoglobinopathies characterized by hemolytic anemia, methemoglobinemia, erythrocytosis and cyanosis.

GLOBIN GENE SWITCHING

The sequential activation and inactivation of globin genes during development shown in Fig. 98-1 is called “hemoglobin switching.” Transcription factors along with epigenetic elements such as DNA methyltransferases and demethylases, interact with enhancers “upstream” of the β -globin gene cluster that contact globin gene promoters, silencing the embryonic and fetal genes. Activation of fetal globin gene repressors during development allows expression of the adult genes. Developmental factors such as RNA-binding factors and microRNAs also impact hemoglobin switching.

β -Globin Gene Switching HbF reactivation by drugs and gene therapy is a prime therapeutic goal for treating the common disorders of hemoglobin, meriting a discussion of the controls of HbF gene silencing. An upstream super-enhancer called the β -globin locus control region (LCR) binds erythroid-specific and ubiquitous transcription factors. The LCR interacts directly with globin gene promoters; transcription factors that silence and activate genes also interact with elements of the globin genes. Competition among the β -like genes for the LCR and autonomous silencing of the embryonic and fetal globin genes depends on transcription factors. Silencing, first of *HBE* and then of *HBG2* and *HBG1*, favors the interaction of the LCR with *HBB*. When *HBG2* or *HBG1* is upregulated by rare point mutations in their promoters, expression of the linked *HBB* is downregulated. Deletions of the *HBB* promoter remove competition for the LCR, increasing the expression of *HBG2*, *HBG1*, and *HBD*. The transcription factors BCL11A (2p16) and ZBTB7A (19p13) silence the HbF genes; BCL11A binds to the HbF gene promoters, repressing them and silencing transcription; ZBTB7A binds upstream of BCL11A with similar

repressive effects. This accounts for the bulk of the switch from HbF to HbA. Mutations in these binding sites abolish the normal silencing of the HbF genes, leading to one type of the benign condition called hereditary persistence of fetal hemoglobin (HPFH). Disruption of the *BCL11A* regulatory elements or the binding sites for BCL11A by gene editing is a prime therapeutic target for HbF induction.

α -Globin Gene Switching A less complex switch takes place in the α -globin gene cluster where a regulatory locus of four elements termed R1-R4 is present within introns of the gene *NPRL3* that is upstream of *HBA2*. A developmental switch from embryonic ζ - to adult α -globin gene expression occurs at about 6 weeks’ gestation.

Modulation of HbF Level Variations in three quantitative trait loci (QTL), *BCL11A*, *MYB* (6q23), and a locus linked to the *HBB* cluster, account for a major portion of HbF variation among normal individuals and patients with sickle cell anemia and β thalassemia. *BCL11A*, a zinc finger protein that represses HbF genes, binds TGACCA motifs, the most important at position -115 in the promoter of each γ -globin gene. ZBTB7A binds 85 nucleotides upstream of these *BCL11A* binding sites; its binding also represses γ -globin gene transcription. When binding of either *BCL11A* or ZBTB7A is disrupted, silencing of *HBG2* and *HBG1* is abrogated. The unique impact of *BCL11A* variants on HbF in sickle cell anemia and β thalassemia is due to their large effect and the high frequency of the variant allele associated with increased HbF.

The *MYB* gene is essential for hematopoiesis and erythroid differentiation. *MYB* inhibits HbF expression directly by activation of *KLF1* and other repressors and indirectly through alteration of the kinetics of erythroid differentiation.

The third QTL is marked by a common variant 158 nucleotides upstream of the transcription start site of *HBG2* and could be a binding site for an uncharacterized HbF repressor. Haplotypes associated with the *HBB* cluster have been defined by single nucleotide polymorphisms (SNPs) among these genes. Sickle cell anemia patients with the Senegal and Arab-Indian HbS gene-associated haplotypes have higher HbF levels than patients with other haplotypes. These two haplotypes have the common-158 C-T variant in the *HBG2* promoter.

DIAGNOSIS OF HEMOGLOBIN DISORDERS

α -Globin gene mutations are expressed in the embryo and fetus and persist throughout life; HbF mutations are expressed in the fetus and in the first months of life, vanishing from notice afterward; δ -globin gene mutations are innocuous and usually not detected; β -globin gene mutations can become clinically apparent after the synthesis of HbF dwindles to stable adult levels.

With rare exceptions, all disorders of hemoglobin are autosomal recessive or co-dominant disorders; a family history of anemia, a common feature of most symptomatic hemoglobinopathies and thalassemias, is often present. In addition to pallor and jaundice, splenomegaly is often present. In sickle cell disease, acute painful vasoocclusive episodes are a diagnostic feature. A small number of laboratory tests can confirm the diagnosis starting with a complete blood count that includes a reticulocyte count with a careful review of a peripheral blood film. A sustained increase in reticulocyte count indicates the presence of hemolytic anemia. Hemoglobin fractionation by high-performance liquid chromatography (HPLC) or capillary electrophoresis, especially when, in addition to the index case, family members are available for study, is often sufficient to confirm a diagnosis at the level of hemoglobin phenotype. DNA sequencing of the globin genes should allow definitive diagnosis. DNA-based diagnosis, which is readily available from excellent reference laboratories, is a prerequisite for most instances of genetic counseling.

Sickle cell disease and β thalassemia have some features in common. They are caused by mutations in the β -globin gene; both are chronic hemolytic anemias sharing complications associated with hemolysis such as venous thrombosis, leg ulcers, and pulmonary hypertension; and they can be cured by hematopoietic stem cell transplantation. Key differences are that only HbS polymerizes and that ineffective

erythropoiesis is a prominent feature of β thalassemia and responsible for its severe anemia. Both diseases could be cured by inducing sufficiently high levels of HbF; in sickle cell disease, HbF prevents the polymerization of HbS; in β thalassemia, sufficient HbF compensates for the deficit of HbA.

SICKLE CELL DISEASE

Sickle cell disease is a clinical and hematologic phenotype caused by an assortment of genotypes (Table 98-2). Sickle cell anemia, defined as homozygosity for the sickle hemoglobin mutation ($\alpha_2\beta^S$; glutamic acid [E] 7 valine [V] GAG-GTG), is the most common of these genotypes, followed by HbSC disease or compound heterozygosity for HbS and HbC ($\alpha_2\beta^C$; E 7 lysine [K] GAG-AAG) genes. Many different thalassemia mutations contribute to the HbS- β thalassemias. The compound heterozygous genotypes are less common than HbS homozygotes; as a rule, their symptoms develop later in life and are less severe. HbS has also been described with many other variant hemoglobins. Few of these genotypes, other than HbSO^{Arab}, HbSE, and HbSD^{Punjab} are symptomatic.

ORIGIN, SPREAD, AND EPIDEMIOLOGY

HbS originated in Africa between 7000 and 22,000 years ago, reaching high frequencies because of the increased genetic fitness of heterozygotes under selective pressure from *Plasmodium falciparum*. The HbS gene became associated with five common β -globin gene haplotypes: Benin, Bantu, Senegal, Cameroon, and Arab-Indian. These haplotypes have a loose association with the severity of disease because each haplotype has a different average level of HbF. In some regions of Africa, India, and the Middle East, nearly half the population have sickle cell trait. Nigeria alone has ~150,000 newborns each year with sickle cell anemia, about one-third of the world's total newborns; most die before age 5. Coerced and free population movement have spread the HbS gene throughout the world. The HbS carrier, or sickle cell trait,

prevalence is 2–15% in emigrant populations; ~100,000 patients in the United States have sickle cell disease; their death in childhood is rare, with the median age of death in the fifth or sixth decade.

PATHOPHYSIOLOGY

Pathophysiologic features of sickle cell disease are summarized in Fig. 98-3. HbS is physiologically similar to HbA in most respects except it polymerizes when deoxygenated. Contacts between one of the β^7 valine residues of deoxyHbS and specific amino acid residues of β - and α -globin culminate in fascicles of hemoglobin that injure the sickle erythrocyte. A delay occurs between the initiation of polymerization and the accumulation of sufficient polymer to damage the cell. It is unclear how much polymer is needed for cell injury, but it is clear that polymer leads directly and indirectly to the multiple abnormalities of the sickle erythrocyte that generate the pathophysiology of disease. Prominent among these abnormalities are HbS polymer penetration of the membrane causing vesiculation with membrane microparticle release; increased activity of the Gardos, K/CL cotransport, and P_{sickle} channels that dehydrate the cell, increasing mean corpuscular sickle hemoglobin concentration (MC[HbS]C), reducing cellular deformability, and increasing the polymerization potential of HbS; translocation of amino phospholipids such as phosphatidylserine to the outer leaflet of the membrane; and oxidation of erythrocyte contents. These and other abnormalities lead to the formation of irreversibly sickled cells (ISCs), which are sickle erythrocytes that are forever deformed because of permanent membrane damage regardless of whether HbS remains polymerized. Damaged sickle erythrocytes are responsible for initiating the vasoocclusive, hemolytic, and inflammatory features of the disease shown in Fig. 98-3.

DIAGNOSIS

Although sickle cell disease can appear in any ethnic group, most often it is present in people of African, Middle Eastern, Mediterranean, and

TABLE 98-2 Common Sickle Hemoglobinopathies

GENOTYPE	CLINICAL ABNORMALITIES	HEMOGLOBIN LEVEL, g/L (g/dL)/MCV, fL	HEMOGLOBIN FRACTIONS (%)
Sickle cell trait (HbAS)	8% of African Americans; hematuria, papillary necrosis, hyposthenuria, increased incidence of chronic kidney disease; 2–4 times increased VTE risk; ? stroke; splenic infarction at altitude; rhabdomyolysis	Normal	HbA: 60–70 HbS: 30–40 Percent HbS dependent on presence or absence of α thalassemia
Sickle cell anemia (HbSS)	Vasoocclusion related: pain, acute chest syndrome, osteonecrosis, splenic infarction Hemolysis related: stroke, pulmonary and systemic vasculopathy, nephropathy, leg ulceration gallstones, priapism, leg ulcers	70–100 (7–10)/80–100	HbS: >75 HbF: 2–25 HbA ₂ : 3–4
HbS- β^0 thalassemia	Rate of complications similar to HbSS	80–100 (8–11)/60–85	HbS: >75 HbF: 2–15 HbA ₂ : 5–6
HbS- β^+ thalassemia	Rate of complications about half the rate of HbSS depending on percent HbA	100–140 (10–14)/70–80	HbS: 60–90 HbA: 5–40 HbF: 1–10 HbA ₂ : 5–6
Hemoglobin SC disease (HbSC)	Nearly asymptomatic to severe disease; about half the rate of complications as HbSS. Increased risk of retinopathy	100–140 (10–14)/70–100	HbS: 50 HbC: 50
HbSE	Resembles clinically HbS- β^+ thalassemia; symptoms delayed; often Asian/Indian ancestry	90–130 (9–13)/65–75	HbS: 65 HbE: 35 HbF: 1–5
HbSS- α thalassemia	Present in 30% of HbSS; phenocopies HbS- β^0 thalassemia; similar to HbSS but with fewer strokes and leg ulcers and less pulmonary vascular and renal disease	80–100 (8–11)/60–85	HbS: >75 HbF: 2–15 HbA ₂ : 4–5
HbS-HPFH	Most common genotype is due to large <i>HBB</i> deletions and is asymptomatic	110–140 (11–14)/70–80	HbS: 70 HbF: 20–30 HbA ₂ : 1–2

Note: Laboratory values are averages in untreated adults.

Abbreviation: VTE, venous thromboembolism.

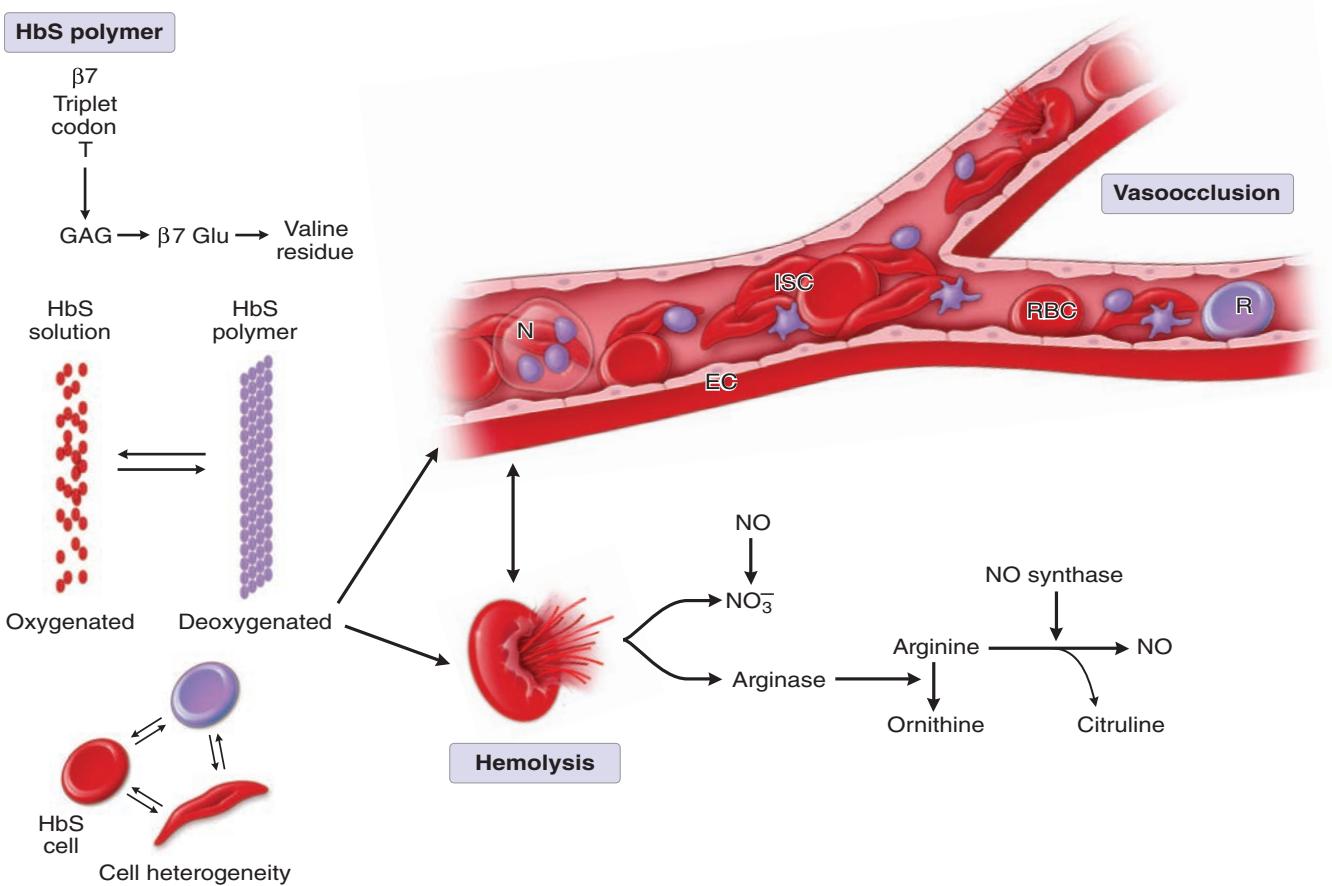


FIGURE 98-3 Pathophysiology of sickle cell disease. HbS is in solution when oxygenated but reversibly polymerizes when deoxygenated. Polymerization is dependent on the 30th power of hemoglobin concentration. In the sickle cell, this means that small changes in hemoglobin concentration or cell hydration can have large effects on polymerization. Polymerization begins seconds to minutes following deoxygenation. Erythrocyte deformation, or sickling, is initially reversible, but after an undetermined number of cell sickling events, the cell becomes irreversibly deformed. These are known as irreversibly sickled cells (ISCs). Their membrane is permanently damaged, although depending on their oxygen (O_2) content, HbS could be in solution. Sickled erythrocytes lead to the clinical and laboratory phenotypes of disease. Sickled cells interact with endothelial cells and other blood cells, occluding flow in small and sometimes large vessels and causing the many complications thought to be a result of vasoocclusion. Sickled cells also live <20 days (normal ~120 days) hemolyzing intra- and extravascularly. Intravascular hemolysis depletes haptoglobin and hemopexin while liberating heme, arginase, and other danger-associated molecular patterns (DAMPs) into the blood. This scavenges nitric oxide (NO), activates platelets and endothelium, reduces antioxidant activity, causes vasoconstriction, and is proinflammatory.

Indian descent. The chief presenting symptom is pain that might be an arthritis-like hand-foot syndrome in young children or the typical acute painful episode in older children and adults. In HbSC disease and HbS-β⁺ thalassemia, acute vasoocclusive episodes occur less often and complications develop later in life; rarely, patients with these genotypes are asymptomatic. The key elements of laboratory diagnosis are outlined in Table 98-2 showing typical hematologic findings and hemoglobin fractions. Figure 98-4 displays HPLC profiles and blood films in typical patients with sickle cell trait, sickle cell anemia, and HbSC disease. Clinical and basic laboratory diagnosis is sufficient for general management and counseling; genetic counseling and family planning usually require DNA-based diagnosis.

■ COMPLICATIONS

Complications of sickle cell disease can be grouped into those that likely are a consequence of sickle vasoocclusion and ones that appeared to be triggered by intravascular hemolysis. Although there is a relationship between these two limbs of pathophysiology, complications associated with vasoocclusion seem to respond best to induction of HbF. Some complications of disease are presented in Table 98-3. Early and effective treatment with hydroxyurea and the integration into management of new treatments discussed below should change this profile.

Acute Painful Episodes Characterized by unprovoked severe pain in extremities or torso that is often symmetrical and stereotypical

for each patient and usually requires treatment with strong opioids in the emergency department, acute painful episodes are the most common acute events in sickle cell disease. They are the chief cause of concern for patients, most of whom have them at some time in their life. Their frequency varies; most patients have one to two episodes a year; some rarely have them; others are hardly ever without them. Acute painful episodes last days to weeks. Complicating the diagnosis and management of the acute pain episode, pain in sickle cell disease can be chronic from complications such as osteonecrosis, osteoporosis, or leg ulcers; chronic and acute pain can overlap; and pain can also be induced by opioid treatment of pain. Diary studies have shown that most of the time patients have some degree of pain that does not reach the intensity of the acute episode. Most patients use oral opioid analgesics for control of this pain. Reliable patients can be given a reasonable supply of oral opioids on a monthly basis.

No diagnostic test can confirm or refute the presence of an acute pain episode; often a 1 to 2 g/dL decrease in hemoglobin level and a modest increase in the leukocyte count are noted during the painful episode. Drastic decreases in hemoglobin and platelet levels with more extreme leukocytosis can portend development of severe acute chest syndrome or multiorgan failure. Acute painful episodes have little to do with the presence of ISCs in the blood or the reticulocyte count. The most anemic patients seem to have the least pain. It is unusual for a cause of acute painful episodes to be identified. Physical examination is not often useful diagnostically. Some patients will have pain on

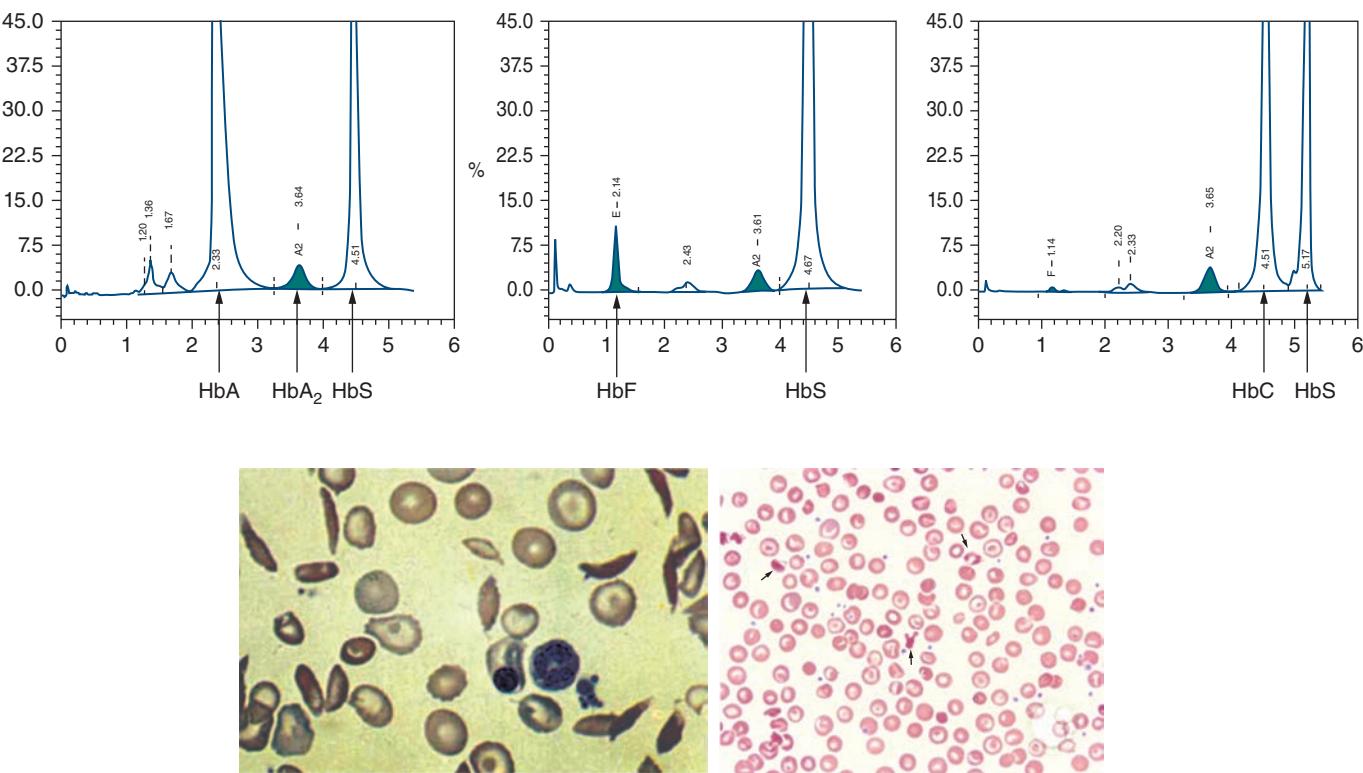


FIGURE 98-4 Diagnosis of sickle cell disease. A. From left to right, high-performance liquid chromatography separation in sickle cell trait, sickle cell anemia, and HbSC disease. Beneath each chromatogram, the individual protein peaks are identified. **B.** Left: Dense, elongated, and pointed cells are the irreversibly sickled cells characteristic of the sickle cell anemia and sickle cell- β^0 thalassemia. Target cells and nucleated red cells are also present. Right: Target cells, cells with squared ends of HbC crystals, and contracted microspherocytes are typical of HbSC disease. (Source: B [right]: Reproduced with permission from American Society of Hematology.)

TABLE 98-3 Complications of Sickle Cell Disease

COMPLICATION	INCIDENCE, DIAGNOSIS, AND FEATURES	TREATMENT
Priapism	~30% of males; can be episodic and short duration (stuttering); severe episodes can cause impotence; associated with markers of hemolysis	Many unproven therapies including α -adrenergic agonists, stilbesterol; consult urology for therapy, which is time-critical
Stroke and silent infarction	10–15% of all cases; infarction in early childhood into adulthood; hemorrhagic in adults; neurocognitive abnormalities in adults even without apparent stroke; associated with markers of hemolysis	Transcranial Doppler screening in children aged 2–16; transfusion for at-risk patients; hydroxyurea
Gallstones/surgery	~40% of patients; bilirubin levels and stones related to polymorphisms of <i>UGT1A</i> ; in surgery requiring general anesthesia, simple preoperative transfusion to a hemoglobin of 10 g/dL is recommended	If asymptomatic, usually let be; otherwise, laparoscopic cholecystectomy
Hepatic disease	>80% of patients have hepatomegaly; intrahepatic cholestasis can have bilirubin ~100 mg/dL; viral hepatitis, iron overload, RBC sequestration, extrahepatic cholestasis also contribute	Exchange transfusion for intrahepatic cholestasis; transplant for end-stage liver failure
Nephropathy	~30% of adults age >30 years; hyperfiltration in children, renal failure in adults; early albuminuria, later nephrotic-range proteinuria; associated with markers of hemolysis	Screen for microalbuminuria by age 10 years; avoid NSAIDs; use ACE inhibitors or receptor antagonists for albuminuria; erythropoietin for symptomatic anemia; dialysis or transplant for renal failure
Lung/pulmonary hypertension	Restrictive disease; asthma common; 5–10% have pulmonary hypertension by right heart catheterization; 30% have increased TRV that portends poor prognosis; associated with markers of hemolysis	Consult expert pulmonologist; screen yearly by echocardiography measurement of TRV
Retinopathy	30% in HbSC disease, 3% in HbSS, ^a develops in peripheral retina; vitreous hemorrhage and retinal detachment can cause blindness	Screen annually starting at age 10 years with fluorescein angiography; laser photocoagulation for proliferative disease
Acute anemic episodes	B19 parvovirus infection, folic acid deficiency, splenic sequestration, delayed hemolytic transfusion reaction with destruction of transfused and sometimes autologous red cells	RBC transfusion if symptomatic; splenectomy if more than one or two episodes of sequestration; anti-parvovirus IgM positive in acute infection, IgG in past infection
Multiorgan failure	Can accompany severe acute chest syndrome; often confused with sepsis and can coexist with sepsis; CNS, liver, muscle, lung, kidney affected	Exchange transfusion, ICU support
Pregnancy	Screening both partners for hemoglobin disorders with risk counseling is critical component of family planning.	All pregnancies are “high risk”; transfuse if sickle cell events increase, if previous miscarriage, multiple fetuses

^aSickle cell anemia (HbSS).

Abbreviations: ACE, angiotensin-converting enzyme; CNS, central nervous system; ICU, intensive care unit; NSAIDs, nonsteroidal anti-inflammatory drugs; TRV, tricuspid regurgitant jet velocity.

pressure over an affected area, perhaps accompanied by swelling; mild fever is common.

Some patients die suddenly shortly after admission for an acute painful episode. The cause of this sudden unexpected death is usually unknown; among the possibilities are arrhythmias and pulmonary embolism. Admitting patients to monitored beds or continuous pulse oximetry for the first 48–72 h of hospitalization might prevent some of these deaths and help identify acute chest syndrome that follows within 72 h in about a quarter of admissions for acute pain. After searching for possible precipitants such as infection or dehydration and treating these appropriately, the foundation of treatment is the proper dosing of opioid analgesics. By the time a patient presents at the emergency department or clinic requesting treatment, they have usually tried nonsteroidal anti-inflammatory drugs (NSAIDs) and oral opioids. In most patients, relief of pain requires the intravenous opioids morphine or hydromorphone. Many patients are opioid tolerant and require higher than usual doses for satisfactory relief. Dosing should not be on an “as-needed” schedule; patient-controlled analgesia or a frequent fixed dose of opioids with rescue doses for breakthrough pain are the preferred means of treatment, with frequent assessments to ensure pain relief without excessive sedation. Adjunctive treatment includes incentive spirometry to forestall pulmonary complications, maintaining hydration with half-normal saline with care not to overhydrate, prophylaxis for thromboembolism, and antihistamines and laxatives to counter expected side effects of opioids; unless hypoxia is present, supplemental O₂ is unnecessary. Ketorolac should not be used, and NSAIDs have little value in patients receiving intravenous opioids.

Acute Chest Syndrome This pneumonia-like illness is the second most frequent acute sickle cell–related event. It occurs in >50% of patients, often more than once. Acute chest syndrome can be mild, especially in children, in whom it can result from viral infection, or devastating, where multiple lobes of the lung are affected with severe hypoxia, multiorgan failure, and death. Chest pain, cough, fever, and hypoxia and a pulmonary infiltrate on chest x-ray are the major diagnostic criteria. The etiology includes in situ thrombosis, emboli, any type of infection, and postoperative hypoventilation. Management in adults is dictated by the severity of the episode. Patients who are hypoxic and febrile are often admitted directly to the intensive care unit. Antibiotics are almost always used in febrile patients even though a causative bacterium is not often cultured. Supplemental O₂ is given for an O₂ saturation <95%. Overhydration and excessive opioids can compound dyspnea and hypoxia. Hypoxic patients who are febrile with leukocytosis and have more than a trivial infiltrate on x-ray are transfused. In the more severely ill patient, exchange transfusion is the preferred modality. When hemoglobin level or symptoms indicate the need for transfusion of the severely ill patient and hours are needed to arrange red cell exchange, simple or top-up transfusion should be started first. Simple transfusions also suffice for less severely affected patients. Most patients survive acute chest syndrome, but in the most severe cases, often caused by embolization of necrotic bone marrow, death can be rapid even with prompt and proper treatment. Thrombocytopenia, leukocyte counts in excess of 20,000/dL, and rapidly developing acute anemia often portend severe acute chest syndrome with the possibility of acute respiratory distress syndrome and multiorgan failure. Many adults have chronic lung disease that could be a sequela of acute chest syndrome, and asthma is very common in patients with sickle cell disease.

Osteonecrosis This painful and sometimes crippling complication that most often affects hips bilaterally occurs in about half of all patients with sickle cell anemia and is also common in HbSC disease; shoulders are less often affected. Beginning with chronic pain that can become severe, loss of function is often the final stage, especially in the hips. MRI can detect the earliest stages, whereas x-ray is less sensitive. Physical therapy and NSAIDs provide some relief; unfortunately, oral opioids are sometimes required. Joint replacement can restore lost mobility and relieve pain, but the life span of prosthetic joints is finite

so surgery should be delayed as long as mobility is satisfactory and pain tolerable.

Leg Ulcers The incidence of leg ulcers is highly dependent on geography and hemoglobin genotype. They are far less common in HbSC disease and HbS-β⁺ thalassemia than in sickle cell anemia and HbS-β⁰ thalassemia. In temperate climates, 10–20% of patients are affected; tropical and subtropical areas have an incidence rate up to 75%; ulcers rarely occur in the Middle East. They can be small and superficial or deep and encompass most of the lower leg. Ulcers can be extraordinarily painful. Long-standing, recurrent large ulcers are difficult to treat. Wet-to-dry dressings and Unna boots are reasonable choices for initial treatment.

SICKLE CELL TRAIT (CARRIERS, OR SIMPLE HETEROZYGOSITY FOR THE HbS GENE)

Carriers of sickle cell trait outnumber patients with the disease by 25 to 1. Counseling and follow-up of carriers detected by cord blood screening are imperfect. Adolescents and adults can forget that they have sickle cell trait. Although usually a benign condition with a normal life expectancy, some features of this trait are shown in Table 98-2. Counseling sickle cell trait carriers about the small risks of complications and their likelihood of having offspring with sickle cell disease is essential. Counseling prior to participation in sports is also important because of the risk, albeit a very small one, of sudden death from heat-related exertional rhabdomyolysis. Optimal hydration before and during exercise can prevent most episodes of heat-related illness.

TREATMENT, SCREENING, COUNSELING, AND ANTENATAL DIAGNOSIS

Patients should, if possible, be referred to a sickle cell center for initial consultation, follow-up, and institution of therapy. Cooperation among primary care providers, hematologists, and other specialists can provide the best preventive care and management of complications. The frequency at which a patient is seen depends on their therapeutic regimen.

Remarkable changes in the treatment landscape have occurred with the promise of even greater benefits from new curative approaches based on gene therapy. The following discussion focuses on treatment to prevent the complications of disease.

Hydroxyurea Hydroxyurea is the standard of care for all patients with sickle cell anemia and HbS-β⁰ thalassemia. It is recommended for patients of all ages regardless of symptoms and should be started in the first year of life. The major mechanism of action of hydroxyurea is to induce high levels of HbF. Hydroxyurea increases HbF unevenly in the red cell population (heterocellularly), so some cells have greater protection from HbS polymerization than others. Although often employed in symptomatic patients with HbSC disease, its benefits in this genotype are understudied. In adults, where the average HbF is ~5%, the increase in HbF is often modest. Nevertheless, pain and acute chest syndrome are reduced by about half, hemoglobin concentration increases by ~1 g/dL, and after 17.5 years of follow-up, mortality was reduced by 49%. In contrast, all children respond robustly to hydroxyurea. When started at <1 year of age at a dose of ~27 mg/kg, HbF levels were 33.3 ± 9.1% and hemoglobin concentration was 10.1 ± 1.3 g/dL. Acute events were markedly reduced with little toxicity. Based on these and other studies in high- and low-resource countries, unless there is a contraindication, hydroxyurea is standard of care for all patients starting in the first year of life at a dose of ~20 mg/kg and titrated to the maximal tolerated dose based on neutrophil and platelet counts.

Voxelotor Voxelotor increases the affinity of the hemoglobin molecule for O₂ (decreases the P₅₀). Voxelotor, 1500 mg daily, was associated with a 1-g/dL increase in hemoglobin concentration in 59% of patients with a reduction in the biomarkers of hemolysis. Although vasoocclusive events were not significantly reduced in the initial report of efficacy, further analysis after a longer observation period suggested that patients achieving the highest hemoglobin had the fewest acute vasoocclusive events. Voxelotor increases hemoglobin-oxygen affinity in all

erythrocytes (pancellularly), and this should provide an increment in polymerization inhibition beyond hydroxyurea. Many questions remain about the long-term effects of voxelotor. Less hemolysis reduces the propensity for stroke, nephropathy, pulmonary hypertension, leg ulcers, and priapism. Will voxelotor be accompanied by these long-term benefits? Could the high O₂ affinity of a modified hemoglobin be harmful for some patients? The answers to these important questions require further study.

Crizanlizumab Downstream effects of HbS polymerization include adhesive interactions among endothelial cells, leukocytes, platelets, and erythrocytes. P-selectin is one molecule involved in these interactions; blocking selectins prevents sickle cell–endothelial adhesion. A P-selectin-blocking monoclonal antibody given intravenously every month reduced acute painful episodes by ~45%, a reduction similar to that seen with hydroxyurea. There were no effects on hemolysis.

L-Glutamine The mechanism of action of this agent, presumed to be the reduction of oxidative stress in sickle erythrocytes, is unsettled. In a phase 3 clinical trial, compared with a placebo, L-glutamine was associated with a 25% reduction in painful episodes and 33% reduction in hospitalization.

There is little consensus regarding how recently approved drugs should be integrated into treatment with hydroxyurea. The effects of voxelotor and crizanlizumab appear to be additive to those of hydroxyurea. Voxelotor can be added to hydroxyurea if the benefits of hydroxyurea alone are insufficient, as they are in most adults. If both hydroxyurea and voxelotor are taken at effective doses and acute vasoocclusive complications continue, crizanlizumab could then be added. The dropout rates in the crizanlizumab and L-glutamine trials was ~35% so adherence to these therapeutics could be problematic.

Transfusion Transfusions are overutilized and underutilized. Major indications for transfusion include severe symptomatic anemia; treatment and prevention of stroke; increasing hemoglobin level to ~10 g/dL before surgery requiring general anesthesia; and acute chest syndrome with hypoxia or multiple lobe involvement. Sometimes transfusions are given during pregnancy when there is a history of complications or fetal loss. Transfusions should usually be avoided in acute pain episodes and for repair of stable chronic anemia. There is a preference for automated red cell exchange transfusion in acute stroke, severe acute chest syndrome, or multiorgan failure or when chronic transfusions are planned. Recent guidelines formulated by experts recommended extended red cell antigen profiling, if possible before the first transfusion, and antigen matching for Rh (C, E or C/c, E/e) and K antigens in addition to ABO/RhD. Complications of transfusion include hyperviscosity, alloimmunization (which occurred in 18.6% of patients transfused between 1979 and 1984 and 27.3% of patients transfused between 2001 and 2011), iron overload, delayed hemolytic transfusion reactions, and hyperhemolysis.

Stem Cell Transplantation Given the excellent results of human leukocyte antigen (HLA)-identical related donor transplants, which have an event-free survival of >95%, this option might be extended to all patients with a suitable donor. Unfortunately, only 15% of patients have a fully matched donor. New approaches to haploidentical transplants are improving event-free survival in these patients.

Preventive Measures and Screening Cord blood screening for sickle cell disease is done in many countries and all 50 states. Affected patients are then directed to clinics that can initiate early preventive care. In childhood, transcranial Doppler screening beginning at age 2 years and repeated annually until age 16 years, prophylactic penicillin (125 mg for children younger than 3 years; 250 mg for children 3 years and older) twice daily until age 5 years, and vaccination with pneumococcal vaccines are the main measures to prevent stroke and invasive pneumococcal infection. Folic acid, 1 mg daily, is given to prevent megaloblastic erythropoiesis; it is probably unnecessary in people with nutritious diets.

All women planning pregnancy should be screened for disorders of hemoglobin by blood counts, erythrocyte indices, and HPLC analysis

of hemoglobin. Individuals with HbS or β thalassemia trait should have their partners tested. Only then is it possible to know the risks of a fetus having sickle cell disease (Table 98-2). Antenatal diagnosis using chorionic villus sampling is widely available.

Emerging Treatments Gene therapy has curative potential and requires neither matched donors nor immunosuppression. Autologous hematopoietic CD34+ stem cells are mobilized and modified ex vivo to produce an antisickling globin. These cells are reinfused following myeloablative conditioning. Phase 1/2 clinical trials have used lentivirus transduction of CD34+ cells with an antisickling β-globin or have interfered with the HbF-suppressive effects of BCL11A using CRISPR/Cas, zinc finger nucleases, or shRNA. These approaches have resulted in HbF or antisickling hemoglobin levels of nearly 50%, reduced hemolysis, total hemoglobin levels of >11 g/dL, and resolution of acute vasoocclusive events. It is too early to know their long-term safety or cure rate.

THALASSEMIA

Thalassemia is caused by reduced accumulation of either α- or β-globin chains causing a relative excess of the unaffected chain. Unbalanced globin synthesis is the hallmark of thalassemia and the proximate cause of its pathophysiology; unpaired globin chains damage the developing erythroblast. Like the HbS mutation and many other red cell traits, thalassemia reached polymorphic levels in tropical and subtropical populations because heterozygotes are protected from *Plasmodium falciparum* infection. Estimates are that 1–5% of the world's population carries a thalassemia mutation; in some locales, most people have a thalassemia mutation. These mutations can affect any globin gene, but clinically, β and α thalassemia are the most important. With nearly 500 unique thalassemia-causing mutations (www.globin.bx.psu.edu) that can interact with each other and with hemoglobinopathies, thalassemia syndromes are remarkably diverse. Where resources permit and the mutation is known, genetic counseling can be provided and antenatal diagnosis is possible.

HbE (β²⁷ glu-lys) is a common variant whose biosynthesis is reduced because the site of the mutation alters its mRNA processing. Its reduced biosynthesis leads to a deficit of β^E-globin chains and features of β thalassemia. Hemoglobin Constant Spring is caused by a mutation of the termination codon of *HBA2* that leads to the synthesis of an elongated α-globin chain that is unstable and suboptimally synthesized. This variant therefore behaves as an α thalassemia variant.

β THALASSEMIA

■ EPIDEMIOLOGY

Once known as Mediterranean anemia, because of the concentration of cases in Italy, Greece, and other countries bordering the Mediterranean Sea, or as Cooley's anemia after the physician first describing cases, β thalassemia is common in most areas of the world where malaria was endemic. Effective programs of screening, counseling, and antenatal diagnosis have reduced the birth of new cases from the Mediterranean region. The bulk of new patients now are of Asian, Middle Eastern, and Indian origin. About 40,000 β thalassemia patients are born yearly. In the United States there are ~1000 cases of severe β thalassemia.

■ CLASSIFICATION

β⁰ Thalassemia mutations totally prevent the accumulation of any globin from the affected gene; β⁺ thalassemia mutations cause minor or extreme reductions in β-globin synthesis. β Thalassemia major and β thalassemia intermedia are now categorized as transfusion-dependent and non-transfusion-dependent based on the number and frequency of transfusions required to sustain a good quality of life.

Pathophysiology Single nucleotide changes are the most common β thalassemia mutations, but gene deletions also occur. A partial listing of the classes of mutations causing β thalassemia include mutations in the promoter elements affecting gene transcription causing mild and sometimes silent β⁺ thalassemia; mutations in the junctions between

exons and introns that affect mRNA processing causing β^0 and β^+ thalassemia; introduction of alternative splice sites into introns or exons usually causing β^+ thalassemia; 3' end-processing sequence mutations preventing RNA polyadenylation leading to mild or silent β^+ thalassemia; mutations preventing initiation of translation causing β^0 thalassemia; and introduction of stop codons that prematurely terminate translation (nonsense mutations) producing reading frameshifts and resulting in truncated globin mRNA and β^0 thalassemia.

In β thalassemia, the deficit in β -globin chain synthesis allows α -globin chains to accumulate in excess. Without a non- α -globin chain partner in dimer and tetramer formation, unpaired α -globin chains are unstable, cannot form a tetramer, and precipitate within the developing erythroblast, causing membrane lipid oxidation and damage. The predominant cause of anemia is intramedullary destruction of erythroid precursors, known as ineffective erythropoiesis. Reduced deformability and phosphatidyl serine exposure also cause extra- and intravascular hemolysis of those erythrocytes that gain entrance into the circulation. In poorly treated β thalassemia, severe anemia leads to bone marrow expansion; hepatosplenomegaly; iron accumulation in liver, heart, and endocrine organs; pulmonary hypertension; and thromboembolic disease.

Frightening pictures of children with severe β thalassemia permeate the literature. These examples of near-terminal disease should be relegated to history because treatment with transfusion and iron chelation can prevent their occurrence and hematopoietic stem cell transplantation can "cure" patients who have suitable donors.

■ DIAGNOSIS

Heterozygous β thalassemia, also known as β thalassemia trait and β thalassemia minor, has mild or no anemia but microcytic/hypochromic erythrocytes with minimal or no increase in reticulocyte count. After recognizing these hematologic abnormalities and excluding iron deficiency, finding an elevated level of HbA₂ and perhaps HbF by HPLC is sufficient to establish this diagnosis. The hematologic characteristics of this heterozygous carrier state are listed in Table 98-4. Sometimes, the spleen is enlarged. Before genetic counseling and antenatal diagnosis are considered after carrier identification by red cell indices and quantitation of HbA₂, the thalassemia-causing mutation should be identified. This is the key to preventing homozygotes or compound heterozygotes with transfusion-dependent thalassemia.

The more severe forms of β thalassemia are hemolytic anemias with hypochromia, microcytosis, reticulocytosis, marked anisocytosis, and

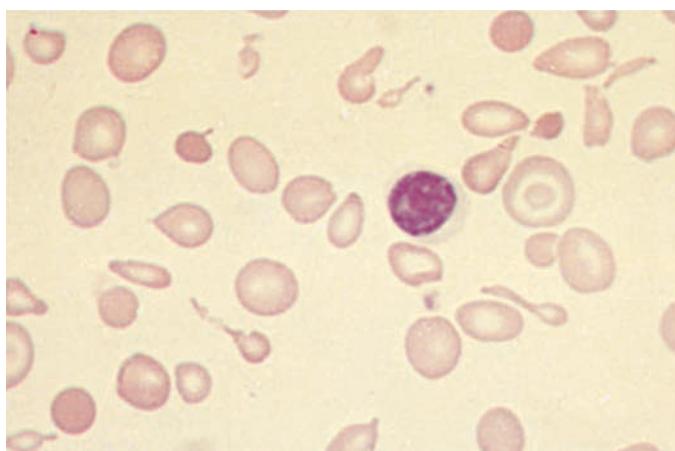


FIGURE 98-5 β Thalassemia intermedia. Target cells and marked variation in cell size and shape but with general hypochromia and microcytosis characterize the blood film. A lymphocyte is shown for size comparison.

poikilocytosis with variable numbers of circulating nucleated red cells (Fig. 98-5).

■ COMPLICATIONS

Complications of severe β thalassemia are many. They are a consequence of chronic hemolytic anemia, chronic transfusion, and iron loading. Increased iron absorption is especially common in non-transfusion-dependent thalassemia. Most complications, listed in Table 98-5, develop because of either inadequate blood transfusion and/or poor iron chelation and iron loading. Even when chelation is optimized, some complications attributable to iron toxicity will develop. Many complications have complex and multifactorial etiologies. Iron stores are estimated by serum ferritin levels; MRI is the most widespread means of noninvasively measuring iron accumulation in liver and heart.

■ MANAGEMENT, SCREENING, COUNSELING, AND ANTENATAL DIAGNOSIS

Heterozygote screening and counseling couples at risk for affected fetuses, with antenatal diagnosis, if needed, is an effective preventive approach. Severe thalassemia should be dealt with in specialized

TABLE 98-4 β Thalassemias

CLASSIFICATION	HEMOGLOBIN (g/dL)/MCV (fL)	HEMOGLOBIN FRACTIONS (%)	CLINICAL FEATURES
β Thalassemia trait	100–140 (10–14)/60–80	HbA: 94 HbF: 1–2 HbA ₂ : 4–6	Heterozygosity for β^+ or β^0 thalassemia mutations; "silent" carriers can have normal HbA ₂ and red cell indices.
Non-transfusion-dependent β thalassemia (thalassemia intermedia)	70–120 (7–12)/65–80	HbA: 60–90 HbF: 10–40 HbA ₂ : 4–6	Defined by infrequent or no transfusion requirement; caused by many different genotypes including homozygosity for "mild" β^+ mutations, combinations of β and α thalassemia, homozygous β thalassemia with high HbF-producing capacity, and many others. Iron loading, thromboembolic disease, and pulmonary hypertension are major clinical events.
Transfusion-dependent β thalassemia (thalassemia major)	20–40 (2–4)/50–80	HbA: 0–5 HbF: 90–100 HbA ₂ : 2–5	Caused by many different genotypes including homozygosity and compound heterozygosity for β^0 and β^+ mutations, combinations of β and α thalassemia; transplantation curative; iron chelation required.
HbE- β thalassemia	50–80 (5–8)/60–70	HbE: 50–70 HbF: 30–50	Common in SE Asian populations; in some parts of the world, the most prevalent severe thalassemia; in HbE- β^0 thalassemia, only HbE and HbF are found; in HbE- β^+ thalassemia, HbA is present. Transfusion dependence depends in part on the thalassemia mutation.
$\delta\beta$ Thalassemia and hemoglobin Lepore	110–120 (11–12)/65–75	HbA: 70 HbF: 7–13 HbA ₂ : 2	Rare; deletions removing the δ - and β -globin genes cause $\delta\beta$ thalassemia; Lepore hemoglobins are fusion globin chains; values are for heterozygotes; homozygotes have 100% HbF with hemoglobin 10–11 g/dL.
Gene deletion HPFH	120–140 (12–14)/75–85	HbA: 70 HbF: 15–30 HbA ₂ : 2	Rare; large deletions removing the δ - and β -globin genes; values are for heterozygotes; homozygotes, who are asymptomatic, have 100% HbF without anemia.

Note: Laboratory results are averages in adults.

TABLE 98-5 Complications of β Thalassemia

COMPLICATION	INCIDENCE, DIAGNOSIS, AND FEATURES
Growth retardation	Most often a feature of delayed or inadequate transfusions but can occur in well-transfused children.
Delayed puberty; secondary amenorrhea	50% and 25%, respectively.
Splenomegaly	Can trap 1–40% of red blood cell volume; increases plasma volume, worsening heart failure. Splenectomy indicated when transfusion requirement to maintain ideal hemoglobin increases. Prophylactic penicillin after splenectomy.
Heart	Due to chronic anemia, heightened sensitivity to iron toxicity, thromboembolic pulmonary hypertension, other causes. Progresses through stages to congestive failure and arrhythmias. Assessed by T2* on MRI. The available chelating agents might have differential effects on different measures of cardiac function and can be used in combination.
Leg ulcers	Common in thalassemia intermedia.
Hepatic disease	Fibrosis progressing to cirrhosis is related to hepatic iron concentration that can be monitored by MRI. Hepatitis also plays a role.
Lung disease/pulmonary hypertension	Fibrosis, chronic thromboembolic disease, restrictive pathophysiology, intravascular hemolysis, and reduced nitric oxide bioavailability
Thromboembolism	Multifactorial etiology including platelet activation, red cell–endothelial interactions, thrombocytosis; endothelial activation; splenectomy.
Endocrinopathies	Diabetes, hypothyroidism, hypoparathyroidism, adrenal insufficiency; hypogonadism; hypothalamic-pituitary axis might be especially sensitive to iron.
Bone disease	Caused by bone marrow expansion, severe iron loading, hypogonadism; osteoporosis in ~50% of patients, even those well treated. Extramedullary hematopoietic masses are a feature of thalassemia intermedia.
Infections	Transfusion associated; linked to iron overload (<i>Yersinia</i>); malaria.

centers where these and other services are available and managed by a team led by a hematologist experienced with this disease with help from endocrinologists, cardiologists, transfusion medicine specialists, and social services.

Transfusion and Iron Chelation Transfusion every 2–4 weeks with a goal pretransfusion hemoglobin concentration of 9–10.5 g/dL, coupled with oral iron chelation to prevent the accumulation of excess toxic iron that accompanies transfusion, has prevented the development of cardiomyopathy and endocrinopathies while extending life to at least 50 years. When to begin transfusions, whether partial exchange transfusion is preferable to simple transfusion, and the choice of blood product require consultation with experts. To be effective, transfusions and iron chelation must be started early, be uninterrupted, and continue lifelong. Older patients who did not have the advantage of effective chelation are more likely to develop multiple disease-related morbidities such as osteoporosis, endocrinopathies, liver disease, and renal failure. Two orally effective chelating agents, deferasirox and deferiprone, and one intravenous chelator, deferoxamine, are available.

Hematopoietic Stem Cell Transplantation There is consensus that patients with available donors should be offered transplantation because of the difficulty of lifelong transfusion and chelation and its imperfect efficacy. Quality of life in successfully transplanted patients exceeds that in patients treated with transfusion and chelation. Transplantation from matched sibling donors is curative in >80% of all cases. Unfortunately, only a third of patients have matched donors. The best results are in the youngest patients who have been effectively chelated and received fewer transfusions. Graft failure, graft rejection,

graft-versus-host disease, and a mortality of 5–20% depending on risk factors are the major drawbacks of this procedure. Results of haploididentical and unrelated donor transplants are improving but lag those of matched sibling donors.

Improving Ineffective Erythropoiesis Luspatercept, a fusion protein containing the extracellular domain of human activin type IIB receptor and the Fc domain of human IgG, was recently approved for treatment of transfusion-dependent thalassemia. By binding transforming growth factor β superfamily ligands and reducing Smad2/3 signaling, luspatercept enhances late-stage erythropoiesis. Given subcutaneously, 1 mg/kg every 3 weeks, it was associated with a 33% reduction in transfusion requirements.

Gene Therapy Lentiviral mediated gene therapy using autologous CD34+ hematopoietic stem cells has been approved in Europe for some patients with transfusion-dependent thalassemia who lack a matched donor. In a clinical trial with a median follow-up of 26 months, where patients received autologous CD34+ cells transduced with a lentiviral vector containing a modified HbA, transfusions were reduced or eliminated and hemoglobin levels stabilized between 8.2 and 13.7 g/dL. However, the results were dependent on the β thalassemia mutation, and although transfusion independence was achieved, some features of disease such as ineffective erythropoiesis were not eliminated. The initial results of CRISPR/Cas editing to downregulate *BCL11A* in β thalassemia have eliminated the need for transfusion and normalized hemoglobin levels (see Sickle Cell Disease).

α THALASSEMIA

In some respects the obverse of β thalassemia, clinically consequential α thalassemia is less common than severe β thalassemia. α Thalassemia is most often found in Asian populations and is usually caused by deletion of α -globin genes rather than point mutations.

EPIDEMIOLOGY

Carriers of the most common α thalassemia chromosomes (Table 98-6) are found in 5–80% of people from tropical and subtropical regions of Africa, the Middle East, India, Southern China, and Melanesia. About 30% of African Americans carry the common $-{\alpha}^7$ chromosome that contains a single functional α -globin gene. HbH disease, the chief clinically important α thalassemia, is most prevalent in Southern China and Southeast Asia. Estimates are that in Thailand ~3500 patients with severe α thalassemia are born yearly. Pregnancies affected by hemoglobin (Hb) Bart's hydrops fetalis occur mainly in Southern China and Southeastern Asia.

CLASSIFICATION

Each normal chromosome 16 contains two α -globin genes; normal diploid individuals have four α -globin genes. A classification of inherited α thalassemia, as summarized in Table 98-6, is based on the number of functional α -globin genes. If one or two α -globin genes are missing or poorly expressed, these people have α thalassemia trait. Their hematologic abnormalities are almost always trivial. HbH disease is usually caused by deletion or malfunction of three α -globin genes. Hb Bart's hydrops fetalis fetuses have no normally functioning α -globin genes. Hundreds of different sized deletions and rarer point mutations affect the production of α globin and the magnitude of imbalanced globin synthesis. Because of this mutational complexity, many different variations of the common α thalassemia syndromes are found.

PATHOPHYSIOLOGY

Reduced accumulation of α -globin leaves non- α -globins unpaired and unable to participate in the formation of functional hemoglobin tetramers. In the fetus, absent or reduced synthesis of α -globin allows unpaired γ -globin chains, which are usually part of the HbF tetramer, to form γ_4 or Hb Bart's; in adults, when γ -globin synthesis is mostly silenced, unpaired β -globin chains, lacking a suitable partner to form HbA, tetramerize as β_4 or HbH. Both Hb Bart's and HbH have very high O₂ affinity and do not unload O₂ in tissues; HbH is also unstable. Severe anemia in Hb Bart's hydrops fetalis is a result of absent normal

TABLE 98-6 α Thalassemias

CLASSIFICATION	α -GLOBIN GENE ARRANGEMENT	HEMOGLOBIN LEVEL, g/L (g/dL)/MCV (fL)	CLINICAL FEATURES
α Thalassemia trait	- $\alpha/\alpha\alpha$ - $\alpha/-\alpha$ - -/ $\alpha\alpha$ $\alpha^{\text{I}}\alpha/\alpha\alpha$	120–150 (12–15)/65–80	The chromosome with one deleted α gene (- α) is called α^{I} thalassemia (α thalassemia-2); the chromosome with both deleted α genes is α^0 thalassemia (α thalassemia-1); non-gene deletion α thalassemias (α^{I}) often have a more severe phenotype.
Hemoglobin H disease	- -/- α $\alpha^{\text{I}}\alpha/-$ $\alpha^{\text{I}}\alpha/\alpha^{\text{I}}\alpha$	50–120 (5–12)/60–70	Mild to moderate anemia depending on genotype; non-gene deletion forms of α thalassemia can produce severe HbH disease.
Hb Bart's hydrops fetalis	--/-		Fatal in utero or at birth with rare survivors. Hydrops can also result from combinations of gene deletion and non-gene deletion α thalassemia.
α Thalassemia/intellectual disability syndromes (ATR-16) (ATR-X)	- -/ $\alpha\alpha$ or - -/- α in ATR-16 $\alpha\alpha/\alpha\alpha$ in ATR-X		ATR-16: Large deletions and rearrangements in chr16p. ATR-X: No α -globin gene deletion or mutation, <i>ATRX</i> mutations, X-linked.
α Thalassemia with myelodysplasia (ATMDS)	$\alpha\alpha/\alpha\alpha$		Mutations in <i>ATRX</i> ; striking male predominance. Hematologic findings of HbH disease.

Note: Laboratory values are averages in adults. $\alpha\alpha/\alpha\alpha$ denotes the chromosome with two intact α -globin genes; - α/α chromosome with one α -globin gene deleted; - -/ $\alpha\alpha$ chromosome with both α -globin genes deleted; α^{I} represents non-gene deletion α thalassemia caused by point mutations. The - α/α chromosome, referred to as α^{I} or α thalassemia-2, most often has a deletion of 3.7 kb of DNA (- $\alpha^{3.7}$) or 4.2 kb of DNA (- $\alpha^{4.2}$) that leaves but a single α -globin gene intact. The chromosome where both α -globin genes are deleted (- -/-) is called α^0 thalassemia or α thalassemia-1. These chromosomes are caused by different-sized deletions that are usually named after their regions of highest frequency such as -SEA, -MED, -FL, and -THAI.

hemoglobin and ineffective erythropoiesis; in HbH disease, unstable HbH leads to oxidative membrane damage with extravascular hemolysis in the spleen and ineffective erythropoiesis.

■ DIAGNOSIS

Microcytosis/hypochromia with nearly normal hemoglobin concentrations, in the absence of iron deficiency and the increased level of HbA₂ that is diagnostic of β thalassemia, is sufficient for a presumptive diagnosis of α thalassemia trait. When genetic counseling is needed and antenatal diagnosis contemplated, the molecular basis of the presumed α thalassemia is required. HbH disease, which is usually due to compound heterozygosity for one chromosome with both α -globin genes deleted and one chromosome with only a single α -globin gene, is defined by the hematologic findings shown in Table 98-6 along with varying levels of reticulocytosis. At birth, when hemoglobin is separated by HPLC, 20–30% Hb Bart's is present; in adults, traces to 40% HbH are present along with residual Hb Bart's in some cases. HbH inclusions can be induced in some red cells after incubation and staining with brilliant cresyl blue. Hemoglobin composition in Hb Bart's hydrops fetalis is predominantly Hb Bart's with some Hb Portland if the deletion removing α -globin genes preserves the ζ -globin gene.

■ COMPLICATIONS

HbH disease is very heterogeneous because of the different combinations of genotypes that can cause this phenotype. Generally, when non-gene deletion mutants, such as Hb Constant Spring, contribute to the genotype, the disease is more severe. In the most common - -/- α genotype, mean hemoglobin in adults is ~11 g/dL Hepatosplenomegaly, jaundice, thalassemic bone changes in the face, and growth impairment are seen 20–50% of cases, depending on the underlying genotype. Iron loading occurs but is not the severe problem it is in β thalassemia. Pregnancy in these patients should be considered high risk and managed accordingly. Mothers of infants with Hb Bart's hydrops fetalis have a history of stillbirth and develop preeclampsia, polyhydramnios, and antepartum hemorrhage and have difficult labor and delivery. Intrauterine transfusion of the fetus is possible.

■ MANAGEMENT, SCREENING, COUNSELING, AND ANTENATAL DIAGNOSIS

When planning families, couples from regions where α thalassemia is common who have red cell indices that suggest the possibility of carrying an α thalassemia gene should have genetic counseling based

on DNA analysis of their globin genes. Iron should be avoided in non-iron-deficient individuals with α thalassemia trait and microcytosis. Transfusions are not usually needed in HbH disease. Nevertheless, depending on the genotype of disease, transfusions might be necessary especially when anemia becomes more severe, for example, with acute anemic episodes or pregnancy. Iron stores should be checked periodically by measuring serum ferritin or MRI; chelation does not appear to be needed.

Hb Bart's hydrops fetalis is best prevented by screening couples at risk and antenatal diagnosis. Intrauterine therapy and perinatal intensive care have permitted survival of some infants with Hb Bart's hydrops fetalis. As growth retardation affects ~40% and neurodevelopmental delay is present in 20% of survivors, prevention is the best approach.

OTHER HEMOGLOBINOPATHIES OF CLINICAL IMPORTANCE (TABLE 98-7)

Thirteen-hundred mutations affecting hemoglobin structure have been described (www.globin.bx.psu.edu). Most are clinically silent. HbC and HbE are common. HbC is found in people of African descent and HbE in South China and Southeast Asia. Heterozygotes for HbC and HbE are clinically well. Even individuals homozygous for these mutations, where the variant hemoglobin comprises >90% of the hemolysate, are clinically well with very mild anemia and microcytosis. The major importance of these variants is the interaction of HbC with HbS and HbE with β thalassemia, as outlined in Tables 98-2 and 98-4. A definitive diagnosis for all rare variants depends on DNA analysis.

Unexpected low O₂ saturation by pulse oximetry (SpO₂) with normal O₂ saturation of arterial blood is occasionally seen in rare hemoglobin variants with clinical phenotypes. Asymptomatic patients with unexpectedly low SpO₂ should not be subjected to unneeded cardio-pulmonary investigations in search of the cause of their "hypoxemia" until the existence of a hemoglobin variant is excluded.

■ M HEMOGLOBINS

M (met) hemoglobins are characterized by oxidation of the heme-iron from its ferrous (Fe⁺⁺) to ferric (Fe⁺⁺⁺) form. The major clinical feature of these disorders is cyanosis that is asymptomatic. Nine M hemoglobin variants have been described. In seven, the mutation involves histidine residues that interact with heme. Asymptomatic slate gray/brownish pseudocyanosis is the main clinical finding. Spectrophotometric recording of the visible spectrum of the hemolysate is

TABLE 98-7 HbC, HbE, and Rare Hemoglobinopathies

CLASSIFICATION	CLINICAL ABNORMALITIES	HEMOGLOBIN LEVEL, g/L (g/dL)/MCV, fL	HEMOGLOBIN FRACTIONS (%)
HbC trait	2% of African Americans; target cells; no disease	Normal	HbC: 30–40 HbA ₂ : 2–3
HbC disease	Target cells; HbC crystals; mild reticulocytosis; splenomegaly	100–130 (10–13)/60–70	HbC: >95 HbF: 2–4 HbA ₂ : 2–3
HbE trait	50% incidence in some Asian populations; a few target cells; clinically normal	120–140 (12–14)/80–90	HbE: 27–31 ^b HbF: 1 HbA ₂ : 3
HbE disease	No hemolysis; 20–80% target cells; no splenomegaly	100–120 (10–12)/65–75	HbE: 85–95 HbF: 3–7 HbA ₂ : 3
High O ₂ affinity hemoglobins	Isolated erythrocytosis; often familial; no splenomegaly; no JAK2 ^{W617F} mutation	150–200 (15–20)	Variants in α- and β-globin genes; patients are heterozygotes; ~25–50% variant
Low O ₂ affinity hemoglobins	Asymptomatic mild anemia; cyanosis	100–140 (10–14)	~50% variant
Unstable hemoglobins	Pigmenturia; hemolysis; reticulocytosis; splenomegaly	90–140 (9–14)/70–90	20–35% variant; rare hyperunstable variants can be undetectable and have the phenotype of thalassemia
M hemoglobins	Some have mild hemolysis; few symptoms	100–140 (10–14)/80–90	20–50% variant depending on gene affected

Note: Laboratory values are averages in adults. As noted for HbAS, the amount of HbC and HbE in heterozygotes depends on the number of α-globin genes.

diagnostic. To distinguish M hemoglobins from methemoglobinemia due to drugs or cytochrome b5 reductase (*CYB5R3*) deficiency, potassium cyanide (KCN) can be added to the hemolysate; methemoglobin-containing blood will turn red, but KCN has no effect on M hemoglobin. Treatment is not needed.

■ UNSTABLE HEMOGLOBINS

Sometimes referred to as congenital Heinz body hemolytic anemias, some mutations result in a hemoglobin tetramer that is unstable and precipitates intracellularly. Such variants are rare and often a result of a new mutation that affects the tertiary or quaternary structure of the molecule. The most common class of mutations introduce a proline residue in the α helix or a polar amino acid into the interior of the molecule. Heinz bodies are intraerythrocytic precipitates that are detectable as dark globular aggregates after staining with a dye such as brilliant cresyl blue. Three unstable hemoglobins are the most common of these rare variants. Hemoglobin Köln (β⁹⁹ val-met) has been found in multiple families, Hb Hasharon (α⁴⁷ asp-his) is found in Ashkenazi Jews, and Hb Zurich (β⁶³ his-arg) is susceptible to oxidant drug-induced hemolysis. Unstable variants present with nonspherocytic hemolytic anemia, but presentation is highly variable. The associated disease is usually mild and does not require transfusion. Heating blood to 50°C or incubation with isopropanol precipitates unstable hemoglobins but must be done with careful controls. Some variants can be detected by HPLC.

■ HEMOGLOBINS WITH HIGH OXYGEN AFFINITY AND LOW OXYGEN AFFINITY

Rare mutations in areas involved in the R-T transition, at critical interfaces between globin chains of the tetramer that reduce the affinity for 2,3-bisphosphoglycerate, or present in the heme pocket account for most of these variants. High O₂ affinity hemoglobins outnumber low O₂ affinity variants by two to one. Isolated erythrocytosis in the absence of splenomegaly suggests the presence of a high O₂ affinity hemoglobin. High O₂ affinity hemoglobin variants shift the hemoglobin-O₂ dissociation curve leftward, causing a low P₅₀ and thereby stimulating erythropoiesis. Many of these variants are due to new mutations. The clinical course is benign, and phlebotomy because of erythrocytosis is usually not required. Early diagnosis is important to forestall unnecessary diagnostic procedures and therapeutics such as cardiac catheterization to exclude congenital heart disease or treatment for polycythemia vera. Low O₂ affinity variants often present with cyanosis. Their hemoglobin-O₂ dissociation curve is right-shifted with

high P₅₀. HPLC might reveal the presence of a hemoglobin variant. Treatment is often not necessary.

■ ACQUIRED DISORDERS OF HEMOGLOBIN

CO binds hemoglobin with high affinity forming carboxyhemoglobin. Carboxyhemoglobin levels can be accurately measured by co-oximetry of arterial blood. Standard pulse oximeters cannot accurately make this measurement. Some newly developed pulse oximeters are able to measure both carboxyhemoglobin and methemoglobin. Bound CO inhibits the transport of O₂; the hemoglobin-O₂ binding curve is left-shifted. Acute and chronic CO intoxication, caused by occupational exposure and other sources of incomplete combustion of hydrocarbons, presents with headache, altered mental status, and other constitutional symptoms. High-flow O₂ via facemask is the preferred treatment; criteria have been developed to guide the use of hyperbaric O₂.

Acquired methemoglobinemia and methemoglobinemia due to deficiency of *CYB5R3* are more common than the M hemoglobins. *CYB5R3* is required for the reduction of methemoglobin by NADH. Affected individuals with “toxic” methemoglobinemia can be cyanotic and symptomatic. As in carboxyhemoglobinemia, O₂ transport is reduced and reflected by the left-shift in the hemoglobin-O₂ binding curve. *CYB5R3* deficiency usually affects only erythrocytes (type I), causing a mild disorder; when all cells are affected (type II), a severe disease results. Intravenous methylene blue is the preferred treatment in symptomatic patients with acquired methemoglobinemia and 40–60% methemoglobin. The usual dose is 1–2 mg/kg. Alternative treatment with ascorbic acid is preferable in people who are glucose-6-phosphate dehydrogenase deficient. Methylene blue interferes with co-oximetry, reducing the value of co-oximetry for monitoring treatment.

Many drugs and chemicals can induce methemoglobin in the absence of *CYB5R3* deficiency. Dapsone and topical anesthetics such as benzocaine are the most common offending agents.

■ FURTHER READING

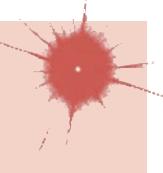
- CHAKRAVORTY S, DICK MC: Antenatal screening for haemoglobinopathies: Current status, barriers and ethics. *Br J Haematol* 187:431, 2019.
- KATO GL et al: Sickle cell disease. *Nat Rev Dis Primers* 4:18010, 2018.
- LEONARD A et al: Curative options for sickle cell disease: Haploididential stem cell transplantation or gene therapy? *Br J Haematol* 189:408, 2020.
- MARKHAM A: Luspatercept: First approval. *Drugs* 80:85, 2020.

- NARDO-MARINO A et al: Emerging therapies in sickle cell disease. *Br J Haematol* 190:149, 2020.
- ORKIN SH, BAUER DE: Emerging genetic therapy for sickle cell disease. *Annu Rev Med* 70:257, 2019.
- PINTO VM et al: Management of the aging beta-thalassemia transfusion-dependent population: The Italian experience. *Blood Rev* 38:100594, 2019.
- STEINBERG MH: Fetal hemoglobin in sickle cell anemia. *Blood* 136:2392, 2020.
- STEINBERG MH: Treating sickle cell anemia: A new era dawns. *Am J Hematol* 95:338, 2020.
- THEIN SL, HOWARD J: How I treat the older adult with sickle cell disease. *Blood* 132:1750, 2018.
- TAHER AT et al: Beta-thalassemia. *N Engl J Med* 384:727, 2021.

99

Megaloblastic Anemias

A. Victor Hoffbrand



The megaloblastic anemias are a group of disorders characterized by the presence of distinctive morphologic appearances of the developing red cells in the bone marrow. The marrow is usually hypercellular, and the anemia is based on ineffective erythropoiesis. The cause is usually a deficiency of either cobalamin (vitamin B_{12}) or folate, but megaloblastic anemia may occur because of genetic or acquired abnormalities that affect the metabolism of these vitamins or because of defects in DNA synthesis not related to cobalamin or folate (Table 99-1).

COBALAMIN

Cobalamin (vitamin B_{12}) exists in a number of different chemical forms. All have a cobalt atom at the center of a corrin ring. In nature, the vitamin is mainly in the 2-deoxyadenosyl (ado) form, which is located in mitochondria. It is the cofactor for the enzyme L-methylmalonyl coenzyme A (CoA) mutase. The other major natural cobalamin is methylcobalamin, the form in human plasma and in cell cytoplasm. It is the cofactor for methionine synthase. Minor amounts of hydroxocobalamin are also present to which methyl- and adocobalamin are converted rapidly by exposure to light.

DIETARY SOURCES AND REQUIREMENTS

Cobalamin is synthesized solely by microorganisms. Ruminants obtain cobalamin from the foregut, but the only source for humans is food of animal origin, for example, meat, fish, and dairy products. Vegetables, fruits, and other foods of nonanimal origin are free from cobalamin unless they are contaminated by bacteria. A normal Western diet contains 5–30 µg of cobalamin daily. Adult daily losses (mainly in the urine and feces) are 1–3 µg (~0.1% of body stores), and because the body

TABLE 99-1 Causes of Megaloblastic Anemia

Cobalamin deficiency or abnormalities of cobalamin metabolism (see Tables 99-3, 99-4)

Folate deficiency or abnormalities of folate metabolism (see Table 99-5)

Therapy with antifolate drugs (e.g., methotrexate)

Independent of either cobalamin or folate deficiency and refractory to cobalamin and folate therapy:

Some cases of acute myeloid leukemia, myelodysplasia

Therapy with drugs interfering with synthesis of DNA (e.g., cytosine arabinoside, hydroxyurea, 6-mercaptopurine, azidothymidine [AZT])

Orotic aciduria (responds to uridine)

Thiamine-responsive

does not have the ability to degrade cobalamin, daily requirements are also about 1–3 µg. Body stores are of the order of 2–3 mg, sufficient for 3–4 years if supplies are completely cut off.

ABSORPTION

Two mechanisms exist for cobalamin absorption. One is passive, occurring equally through buccal, duodenal, and ileal mucosa; it is rapid but extremely inefficient, with <1% of an oral dose being absorbed by this process. The normal physiologic mechanism is active; it occurs through the ileum and is efficient for small (a few micrograms) oral doses of cobalamin, and it is mediated by gastric intrinsic factor (IF). Dietary cobalamin is released from protein complexes by enzymes in the stomach, duodenum, and jejunum; it combines rapidly with a salivary glycoprotein that belongs to the family of cobalamin-binding proteins known as haptocorrins (HCs). In the intestine, the HC is digested by pancreatic trypsin and the cobalamin is transferred to IF.

IF (gene at chromosome 11q13) is produced in the gastric parietal cells of the fundus and body of the stomach, and its secretion parallels that of hydrochloric acid. Normally, a vast excess of IF is available. The IF-cobalamin complex passes to the ileum, where IF attaches to a specific receptor (cubilin) on the microvillus membrane of the enterocytes. Cubilin also is present in yolk sac and renal proximal tubular epithelium. Cubilin appears to traffic by means of amnionless (AMN), an endocytic receptor protein that directs sublocalization and endocytosis of cubilin with its ligand IF-cobalamin complex. The cobalamin-IF complex enters the ileal cell, where IF is destroyed. After a delay of about 6 h, the cobalamin appears in portal blood attached to transcobalamin (TC) II.

Between 0.5 and 5 µg of cobalamin enter the bile each day. This binds to IF, and a major portion of biliary cobalamin normally is reabsorbed together with cobalamin derived from sloughed intestinal cells. Because of the appreciable amount of cobalamin undergoing enterohepatic circulation, cobalamin deficiency develops more rapidly in individuals who malabsorb cobalamin than it does in vegans, in whom reabsorption of biliary cobalamin is intact.

TRANSPORT

Two main cobalamin transport proteins exist in human plasma; they both bind cobalamin—one molecule for one molecule. One HC, also known as TC I, is closely related to other cobalamin-binding HCs in milk, gastric juice, bile, saliva, and other fluids. The gene *TCNL* is at chromosome 11q11-q12.3. These HCs differ from each other only in the carbohydrate moiety of the molecule. TC I is derived primarily from the specific granules in neutrophils. Normally, it is about two-thirds saturated with cobalamin, which it binds tightly. TC I does not enhance cobalamin entry into tissues. Glycoprotein receptors on liver cells are involved in the removal of TC I from plasma, and TC I may play a role in the transport of cobalamin analogues (which it binds more effectively than IF) to the liver for excretion in bile.

The other major cobalamin transport protein in plasma is transcobalamin, also known as TC II. The gene is on chromosome 22q11-q13.1. As for IF and HC, there are nine exons. The three proteins are likely to have a common ancestral origin. TC II is synthesized by liver and by other tissues, including macrophages, ileum, and vascular endothelium. It normally carries only 20–60 ng of cobalamin per liter of plasma and readily gives up cobalamin to marrow, placenta, and other tissues, which it enters by receptor-mediated endocytosis involving the TC II receptor and megalin (encoded by the *LRP-2* gene). The TC II cobalamin is internalized by endocytosis via clathrin-coated pits; the complex is degraded, but the receptor probably is recycled to the cell membrane as is the case for transferrin. Export of “free” cobalamin is via the ATP-binding cassette drug transporter alias multidrug resistance protein 1.

FOLATE

DIETARY FOLATE

Folic (pteroylglutamic) acid is a yellow, crystalline, water-soluble substance. It is the parent compound of a large family of natural folate compounds, which differ from it in three respects: (1) they are partly or

TABLE 99-2 Biochemical Reactions of Folate Coenzymes

REACTION	COENZYME FORM OF FOLATE INVOLVED	SINGLE CARBON UNIT TRANSFERRED	IMPORTANCE
<i>Formate activation</i>	THF	-CHO	Generation of 10-formyl-THF
<i>Purine synthesis</i>	Formation of glycinamide ribonucleotide	5,10-Methylene-THF	Formation of purines needed for DNA, RNA synthesis, but reactions probably not rate-limiting
	Formylation of aminoimidazole carboxamide ribonucleotide (AICAR)	10-Formyl (CHO)THF	
<i>Pyrimidine synthesis</i>	Methylation of deoxyuridine monophosphate (dUMP) to thymidine monophosphate (dTMP)	5,10-Methylene-THF	Rate limiting in DNA synthesis Oxidizes THF to DHF Some breakdown of folate at the C-9–N-10 bond
<i>Amino acid interconversion</i>	Serine-glycine interconversion	THF	Entry of single carbon units into active pool Demethylation of 5-MTHF to THF; also requires cobalamin, flavine adenine dinucleotide, ATP, and adenosylmethionine
	Homocysteine to methionine	5-Methyl(M)THF	
	Forminoglutamic acid to glutamic acid in histidine catabolism	THF	

Abbreviations: DHF, dihydrofolate; THF, tetrahydrofolate.

completely reduced to dihydrofolate (DHF) or tetrahydrofolate (THF) derivatives, (2) they usually contain a single carbon unit (Table 99-2), and (3) 70–90% of natural folates are folate-polyglutamates.

Most foods contain some folate. The highest concentrations are found in liver, yeast, spinach, other greens, and nuts (>100 µg/100 g). The total folate content of an average Western diet is ~250 µg daily, but the amount varies widely according to the type of food eaten and the method of cooking. Folate is easily destroyed by heating, particularly in large volumes of water. Total-body folate in the adult is ~10 mg, with the liver containing the largest store. Daily adult requirements are ~100 µg, and so stores are sufficient for only 3–4 months in normal adults, and severe folate deficiency may develop rapidly.

■ ABSORPTION

Folates are absorbed rapidly from the upper small intestine. The absorption of folate polyglutamates is less efficient than that of monoglutamates; on average, ~50% of food folate is absorbed. Polyglutamate forms are hydrolyzed to the monoglutamate derivatives either in the lumen of the intestine or within the mucosa. All dietary folates are converted to 5-methyl-THF (5-MTHF) within the small intestinal mucosa before entering portal plasma. The monoglutamates are actively transported across the enterocyte by a proton-coupled folate transporter (PCFT, SCL46A1). This is situated at the apical brush border and is most active at pH 5.5, which is about the pH of the duodenal and jejunal surface. Genetic mutations of this protein underlie hereditary malabsorption of folate (see below). Pteroylglutamic acid at doses >400 µg is absorbed largely unchanged and converted to natural folates in the liver. Lower doses are converted to 5-MTHF during absorption through the intestine.

About 60–90 µg of folate enter the bile each day and are excreted into the small intestine. Loss of this folate, together with the folate of sloughed intestinal cells, accelerates the speed with which folate deficiency develops in malabsorption conditions.

■ TRANSPORT

Folate is transported in plasma; about one-third is loosely bound to albumin, and two-thirds are unbound. In all body fluids (plasma, cerebrospinal fluid, milk, bile), folate is largely, if not entirely, 5-MTHF in the monoglutamate form. Three types of folate-binding protein are involved. A reduced folate transporter (RFC, SLC19A1) is the major route of delivery of plasma folate (5-MTHF) to cells. Two folate receptors, FR2 and FR3 embedded in the cell membrane by a glycosyl phosphatidylinositol anchor, transport folate into the cell via receptor-mediated endocytosis. The third protein, proton-coupled folate

transporter (PCFT), transports folate at low pH from the vesicle to the cell cytoplasm. The reduced folate transporter also mediates uptake of methotrexate by cells.

■ BIOCHEMICAL FUNCTIONS

Folates (as the intracellular polyglutamate derivatives) act as coenzymes in the transfer of single-carbon units (Fig. 99-1 and Table 99-2). Two of these reactions are involved in purine synthesis and one in pyrimidine synthesis necessary for DNA and RNA replication. Folate is also a coenzyme for methionine synthesis, in which methylcobalamin is also involved and in which THF is regenerated. THF is the acceptor of single carbon units newly entering the active pool via conversion of serine to glycine. Methionine, the other product of the methionine synthase reaction, is the precursor for S-adenosylmethionine (SAM), the universal methyl donor involved in >100 methyltransferase reactions (Fig. 99-1).

During thymidylate synthesis, 5,10-methylene-THF is oxidized to DHF. The enzyme DHF reductase converts this to THF. The drugs methotrexate, pyrimethamine, and (mainly in bacteria) trimethoprim inhibit DHF reductase and so prevent formation of active THF coenzymes from DHF. A small fraction of the folate coenzyme is not recycled during thymidylate synthesis but is degraded at the C9-N10 bond.

BIOCHEMICAL BASIS OF MEGALOBLASTIC ANEMIA

The common feature of all megaloblastic anemias is a defect in DNA synthesis that affects rapidly dividing cells in the bone marrow. All conditions that give rise to megaloblastic changes have in common a disparity in the rate of synthesis or availability of the four immediate precursors of DNA: the deoxyribonucleoside triphosphates (dNTPs)—dA(adenine)TP and dG(guanine)TP (purines), dT(thymine)TP, and dC(cytosine)TP (pyrimidines). In deficiencies of either folate or cobalamin, there is failure to convert deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTDP), the precursor of dTTP (Fig. 99-1). This is the case because folate is needed as the coenzyme 5,10-methylene-THF polyglutamate for conversion of dUMP to dTDP; the availability of 5,10-methylene-THF is reduced in either cobalamin or folate deficiency. DNA replication from multiple origins along the chromosome is slower than normal during mitosis, and there is failure of joining up the incomplete replicons with resulting single-stranded DNA breaks. An alternative theory for megaloblastic anemia in cobalamin or folate deficiency is misincorporation of uracil into DNA because of the accumulation of deoxyuridine triphosphate (dUTP) at

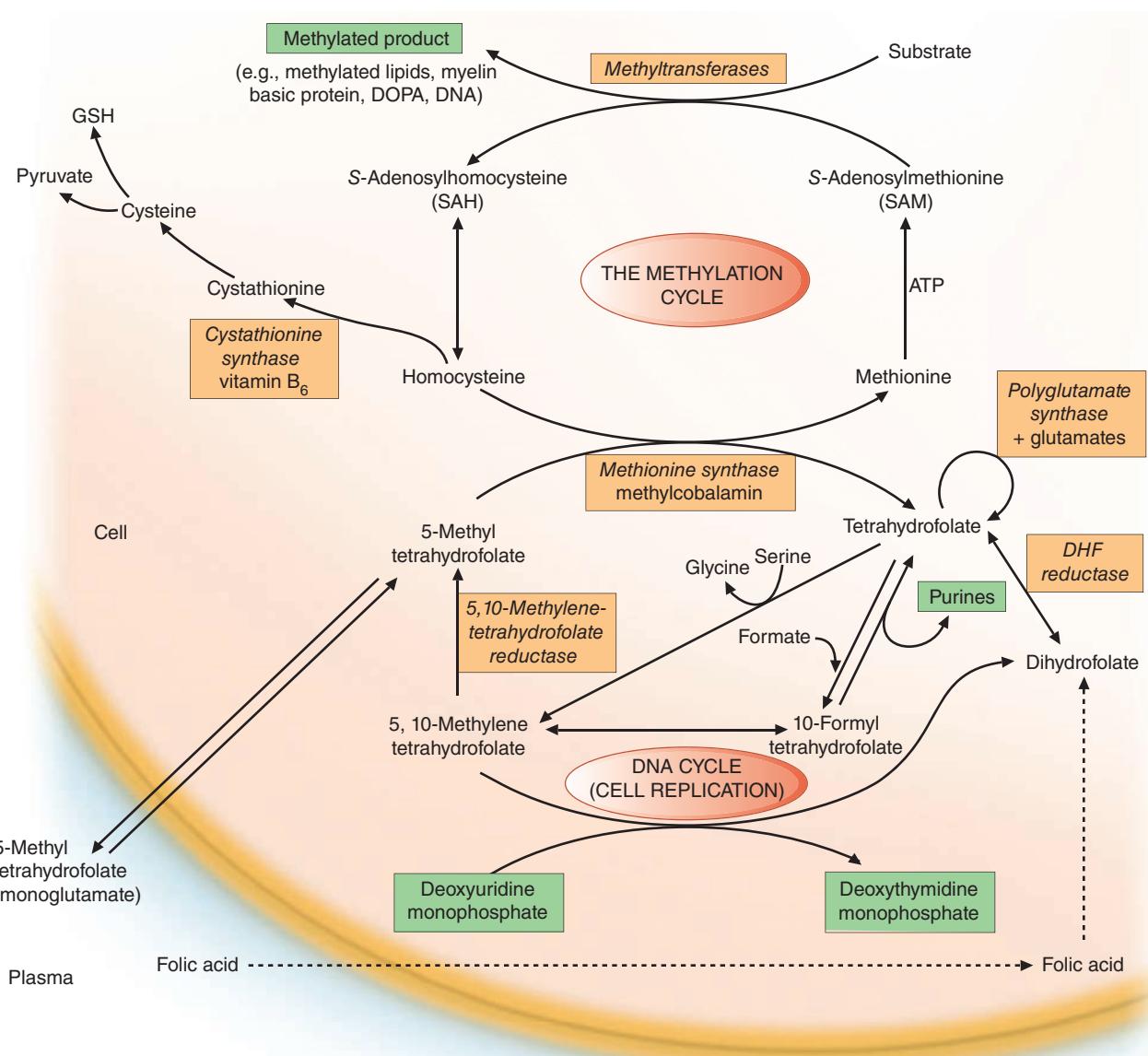


FIGURE 99-1 The role of folates in DNA synthesis and in formation of S-adenosylmethionine (SAM), which is involved in numerous methylation reactions. DHF, dihydrofolate; GSH, glutathione. (Reproduced with permission from AV Hoffbrand et al [eds]: Postgraduate Haematology, 5th ed. Oxford, UK, Blackwell Publishing, 2005.)

the DNA replication fork as a consequence of the block in conversion of dUMP to dTMP.

COBALAMIN-FOLATE RELATIONS

Folate is required for many reactions in mammalian tissues. Only two reactions in the body are known to require cobalamin. Methylmalonyl-CoA isomerization requires adocobalamin, and the methylation of homocysteine to methionine requires both methylcobalamin and 5-MTHF (Fig. 99-1). This reaction is the first step in the pathway by which 5-MTHF, which enters bone marrow and other cells from plasma, is converted into all the intracellular folate coenzymes. The coenzymes are all polyglutamated (the larger size aiding retention in the cell), but the enzyme folate polyglutamate synthase can use only THF, not MTHF, as substrate. In cobalamin deficiency, MTHF accumulates in plasma, and intracellular folate concentrations fall due to failure of formation of THF, the substrate on which folate polyglutamates are built. This has been termed *THF starvation*, or the *methylfolate trap*.

This theory explains the abnormalities of folate metabolism that occur in cobalamin deficiency (high serum folate, low cell folate, positive purine precursor aminoimidazole carboxamide ribonucleotide [AICAR] excretion; Table 99-2) and also why the anemia of cobalamin deficiency responds to folic acid in large doses.

CLINICAL FEATURES

Many symptomless patients are detected through the finding of a raised mean corpuscular volume (MCV) on a routine blood count. The main clinical features in more severe cases are those of anemia. Anorexia is usually marked, and there may be weight loss, diarrhea, or constipation. Glossitis, angular cheilosis, a mild fever in more severely anemic patients, jaundice (unconjugated), and reversible melanin skin hyperpigmentation also may occur with a deficiency of either folate or cobalamin. Thrombocytopenia sometimes leads to bruising, and this may be aggravated by vitamin C deficiency or alcohol in malnourished patients. The anemia and low leukocyte count may predispose to infections, particularly of the respiratory and urinary tracts. Cobalamin deficiency has also been associated in a few studies with impaired bactericidal function of phagocytes and with osteoporosis.

Neurologic Manifestations Vitamin B₁₂ is needed for the myelination of the central nervous system. Its deficiency may cause a bilateral peripheral neuropathy or degeneration (demyelination) of the cervical and thoracic posterior and lateral (pyramidal) tracts of the spinal cord and, less frequently, of the cranial nerves and of the white matter of the brain. Optic atrophy and cerebral symptoms including dementia, depression, psychotic symptoms, and cognitive impairment may be

prominent. There may also be anosmia and loss of taste. MRI may show the “spongy” degeneration of the cord.

The patient, more frequently male, typically presents with paresthesias, muscle weakness, or difficulty in walking but sometimes may present with dementia, psychotic disturbances, or visual impairment. There is usually loss of proprioception and vibration sensation with positive Romberg and Lhermitte signs. Gait may be ataxic with spasticity (hyperreflexia). Autonomic nervous dysfunction can result in postural hypotension, impotence, and incontinence.

Long-term nutritional cobalamin deficiency in infancy leads to poor brain development and impaired intellectual development. In infancy, there may be feeding difficulties, lethargy, and coma. Convulsions and myoclonus have been described. An important clinical problem is the nonanemic patient with neurologic or psychiatric abnormalities and a low or borderline serum cobalamin level. In such patients, it is necessary to try to establish whether there is significant cobalamin deficiency, for example, by careful examination of the blood film, tests for pernicious anemia (PA) by serum gastrin level and for antibodies to IF or parietal cells, along with serum methylmalonic acid (MMA) measurement if available. A trial of cobalamin therapy for at least 3 months will usually also be needed to determine whether the symptoms improve.

The biochemical basis for cobalamin neuropathy remains obscure. Its occurrence in the absence of methylmalonic aciduria in TC II deficiency suggests that the neuropathy is related to the defect in homocysteine-methionine conversion. Accumulation of S-adenosylhomocysteine in the brain, resulting in inhibition of transmethylation reactions, has been suggested. Folate deficiency has been suggested to cause organic nervous disease, but this is uncertain, although methotrexate injected into the cerebrospinal fluid may cause brain or spinal cord damage.

Psychiatric disturbance as discussed above is common in both folate and cobalamin deficiencies. This, like the neuropathy, has been attributed to a failure of the synthesis of SAM, which is needed in methylation of biogenic amines (e.g., dopamine) as well as that of proteins, phospholipids, and neurotransmitters in the brain (Fig. 99-1). Associations between lower serum folate or cobalamin levels and higher homocysteine levels and the development of decreased cognitive function and dementia in Alzheimer's disease have been reported. A meta-analysis of randomized, placebo-controlled trials of homocysteine-lowering B-vitamin supplementation of individuals with and without cognitive impairment, however, showed that supplementation with vitamin B₁₂, vitamin B₆, and folic acid alone or in combination did not improve cognitive function. Some studies done in China suggest some cognitive improvement with supplements of both vitamins. It is unknown whether prolonged treatment with these B vitamins can reduce the risk of dementia in later life.

■ GENERAL TISSUE EFFECTS OF COBALAMIN AND FOLATE DEFICIENCIES

Epithelial Surfaces After the marrow, the next most frequently affected tissues are the epithelial cell surfaces of the mouth (with glossitis), stomach, and small intestine and the respiratory, urinary, and female genital tracts. The cells show macrocytosis, with increased numbers of multinucleate and dying cells. The deficiencies may cause cervical smear abnormalities.

Complications of Pregnancy The gonads are also affected, and infertility is common in both men and women with severe deficiency of either vitamin. Maternal folate deficiency has been implicated as a cause of prematurity, and both folate deficiency and cobalamin deficiency have been implicated in recurrent fetal loss and neural tube defects, as discussed below.

Neural Tube Defects Folic acid supplements at the time of conception and in the first 12 weeks of pregnancy reduce by ~70% the incidence of neural tube defects (NTDs) (anencephaly, meningomyelocele, encephalocele, and spina bifida) in the fetus. Most of this protective effect can be achieved by taking folic acid, 0.4 mg daily, at the time of conception.

The incidence of cleft palate and harelip also can be reduced by prophylactic folic acid. There is no clear simple relationship between maternal folate status and these fetal abnormalities, although overall, the lower the maternal folate, the greater is the risk to the fetus. NTDs also can be caused by antifolate and antiepileptic drugs.

An underlying maternal folate metabolic abnormality has also been postulated. One abnormality has been identified: reduced activity of the enzyme 5,10-methylene-THF reductase (MTHFR) (Fig. 99-1) caused by a common C677T polymorphism in the *MTHFR* gene. In one study, the prevalence of this polymorphism was found to be higher than in controls in the parents of NTD fetuses and in the fetuses themselves: homozygosity for the TT mutation was found in 13% of cases compared with 5% of control subjects. The polymorphism codes for a thermolabile form of MTHFR. The homozygous state results in a lower mean serum and red cell folate level compared with control subjects, as well as significantly higher serum homocysteine levels. Tests for mutations in other enzymes possibly associated with NTDs, for example, methionine synthase and serine-glycine hydroxymethylase, have been negative. Serum vitamin B₁₂ levels are also lower in the sera of mothers of NTD infants than in controls. In addition, maternal TC II receptor polymorphisms are associated with increased risk of NTD births. However, no studies show that dietary fortification with vitamin B₁₂ reduces the incidence of NTDs.

Cardiovascular Disease Children with severe homocystinuria (blood levels $\geq 100 \mu\text{mol/L}$) due to deficiency of one of three enzymes (methionine synthase, MTHFR, or cystathione synthase; Fig. 99-1) have vascular disease, for example, ischemic heart disease, cerebrovascular disease, or pulmonary embolus, as teenagers or in young adulthood. Lesser degrees of raised serum homocysteine and low levels of serum folate and homozygous inherited mutations of *MTHFR* have been found to be associated with cerebrovascular, peripheral vascular, and coronary heart disease and with deep vein thrombosis. Prospective randomized trials of lowering homocysteine levels with supplements of folic acid, vitamin B₁₂, and vitamin B₆ against placebo over a 5-year period in patients with vascular disease or diabetes have not, however, shown a reduction of first event fatal or nonfatal myocardial infarction, nor have these supplements reduced the risk of recurrent cardiovascular disease after an acute myocardial infarct. Meta-analysis showed an 18% reduction in strokes. The benefit for stroke prevention has been confirmed by a large ($>20,000$ subjects) randomized prospective study in hypertensive subjects in China. This showed a significant reduction in the first incidence of stroke in subjects receiving enalapril and folic acid compared to enalapril alone. The effect was especially marked in the subjects commencing the prospective trial with the lowest serum folate levels. Venous thrombosis has been reported to be more frequent in folate-deficient or vitamin B₁₂-deficient subjects than in controls and to occur at unusual sites such as cerebral venous sinuses. This tendency was ascribed to raised plasma homocysteine levels in folate or vitamin B₁₂ deficiency.

Malignancy Prophylactic folic acid in pregnancy has been found in some but not all studies to reduce the subsequent incidence of acute lymphoblastic leukemia (ALL) in childhood. A significant negative association has also been found with the *MTHFR* C677T polymorphism and leukemias with mixed lineage leukemia (MLL) translocations, but a positive association was found with hyperdiploidy in infants with ALL or acute myeloid leukemia or with childhood ALL. A second polymorphism in the *MTHFR* gene, A1298C, is also strongly associated with hyperdiploid leukemia. Various positive and negative associations are noted between polymorphisms in folate-dependent enzymes and the incidence of adult ALL. The C677T polymorphism is thought to lead to increased thymidine pools and “better quality” of DNA synthesis by shunting one-carbon groups toward thymidine and purine synthesis. This may explain its reported association with a lower risk for colorectal cancer. Most but not all studies suggest that prophylactic folic acid also protects against colon adenomas. Other tumors that have been associated with folate polymorphisms or status include follicular lymphoma, breast cancer, and gastric cancer. A meta-analysis of 50,000 individuals given folic acid (0.5–40 mg daily) or placebo in cardiovascular or colon

adenoma prevention trials found that folic acid supplementation did not significantly increase or decrease the overall incidence of cancer or of any site-specific cancer during a weighted average scheduled treatment duration of 5.7 years. Because folic acid may "feed" tumors, it probably should be avoided in those with established tumors unless there is severe megaloblastic anemia due to folate deficiency.

HEMATOLOGIC FINDINGS

PERIPHERAL BLOOD

Oval macrocytes, usually with considerable anisocytosis and poikilocytosis, are the main feature (Fig. 99-2A). The MCV is usually >100 fL unless a cause of microcytosis (e.g., iron deficiency or thalassemia trait) is present. Some of the neutrophils are hypersegmented (more than five nuclear lobes). There may be leukopenia due to a reduction in granulocytes and lymphocytes, but this is usually $>1.5 \times 10^9/L$; the platelet count may be moderately reduced, rarely to $<40 \times 10^9/L$. The severity of all these changes parallels the degree of anemia. In a nonanemic patient, the presence of a few macrocytes and hypersegmented neutrophils in the peripheral blood may be the only indication of the underlying disorder.

BONE MARROW

In a severely anemic patient, the marrow is hypercellular with an accumulation of primitive cells due to selective death by apoptosis of more mature forms. The erythroblast nucleus maintains a primitive appearance despite maturation and hemoglobinization of the cytoplasm. The cells are larger than normoblasts, and an increased number of cells with eccentric lobulated nuclei or nuclear fragments may be present (Fig. 99-2B). Giant and abnormally shaped metamyelocytes and enlarged hyperpolyploid megakaryocytes are characteristic. In severe cases, the accumulation of primitive cells may mimic acute myeloid leukemia, whereas in less anemic patients, the changes in the marrow may be difficult to recognize. The terms *intermediate*, *mild*, and *early* have been used. The term *megaloblastoid* does not mean mildly megaloblastic. It is used to describe cells with both immature-appearing nuclei and defective hemoglobinization and is usually seen in myelodysplasia.

CHROMOSOMES

Bone marrow cells, transformed lymphocytes, and other proliferating cells in the body show a variety of changes, including random breaks, reduced contraction, spreading of the centromere, and exaggeration of

secondary chromosomal constrictions and overprominent satellites. Similar abnormalities may be produced by antimetabolite drugs (e.g., cytosine arabinoside, hydroxyurea, and methotrexate) that interfere with either DNA replication or folate metabolism and that also cause megaloblastic appearances.

INEFFECTIVE HEMATOPOIESIS

Unconjugated bilirubin accumulates in plasma due to the death of nucleated red cells in the marrow (ineffective erythropoiesis). Other evidence for this includes raised urine urobilinogen, reduced haptoglobins and positive urine hemosiderin, and a raised serum lactate dehydrogenase. A weakly positive direct antiglobulin test due to complement can lead to a false diagnosis of autoimmune hemolytic anemia.

CAUSES OF COBALAMIN DEFICIENCY

Cobalamin deficiency is usually due to malabsorption. The only other cause is inadequate dietary intake.

INADEQUATE DIETARY INTAKE

Adults Dietary cobalamin deficiency arises in vegans who omit meat, fish, eggs, cheese, and other animal products from their diet. The largest group in the world consists of Hindus, and it is likely that many millions of Indians are at risk of deficiency of cobalamin on a nutritional basis. Subnormal serum cobalamin levels are found in up to 50% of randomly selected, young, adult Indian vegans, but the deficiency usually does not progress to megaloblastic anemia since the diet of most vegans is not totally lacking in cobalamin and the enterohepatic circulation of cobalamin is intact. Dietary cobalamin deficiency may also arise rarely in nonvegetarian individuals who exist on grossly inadequate diets because of poverty or psychiatric disturbance.

Infants Cobalamin deficiency has been described in infants born to severely cobalamin-deficient mothers. These infants develop megaloblastic anemia at about 3–6 months of age, presumably because they are born with low stores of cobalamin and because they are fed breast milk with low cobalamin content. The babies have also shown growth retardation, impaired psychomotor development, and other neurologic sequelae. MRI shows delayed myelination and atrophy.

GASTRIC CAUSES OF COBALAMIN MALABSORPTION

See Tables 99-3 and 99-4.

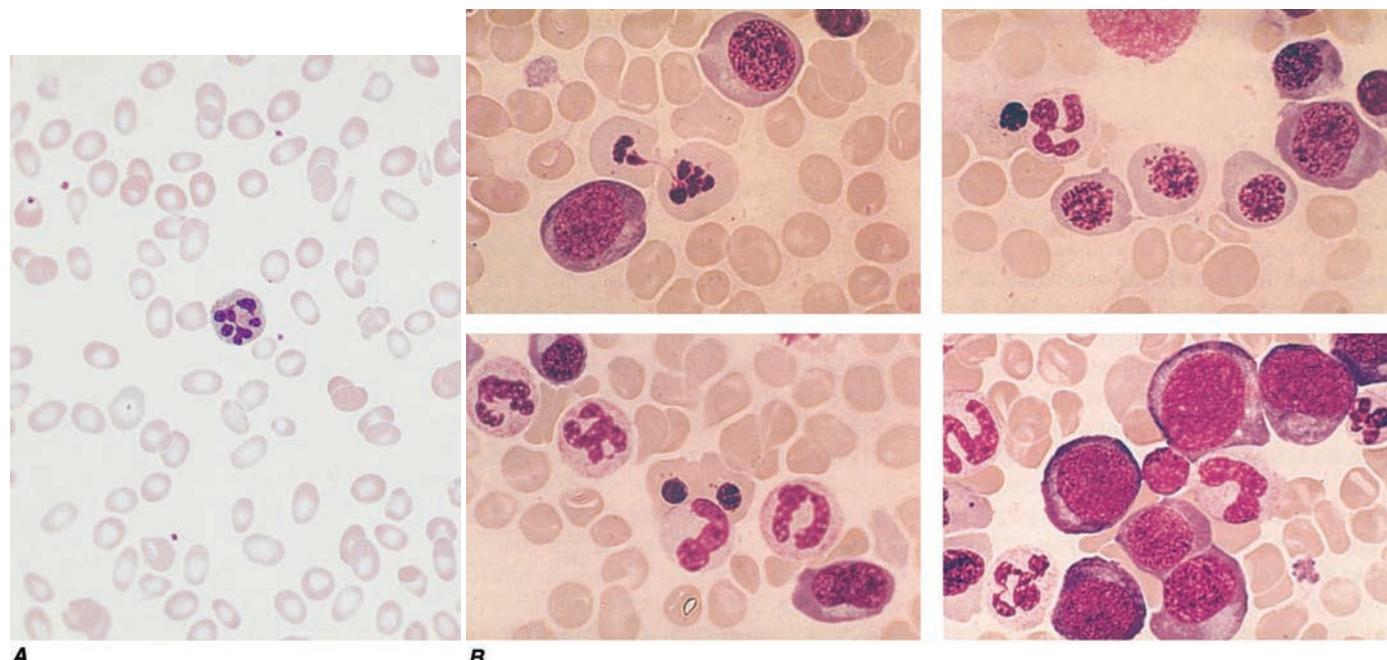


FIGURE 99-2 **A.** The peripheral blood in severe megaloblastic anemia. **B.** The bone marrow in severe megaloblastic anemia. (Reprinted from AV Hoffbrand et al [eds]: Postgraduate Haematology, 5th ed. Oxford, UK, Blackwell Publishing, 2005; with permission.)

TABLE 99-3 Causes of Cobalamin Deficiency Sufficiently Severe to Cause Megaloblastic Anemia

NUTRITIONAL	VEGANS
Malabsorption	Pernicious anemia
Gastric causes	Congenital absence of intrinsic factor or functional abnormality
	Total or partial gastrectomy
Intestinal causes	Intestinal stagnant loop syndrome: jejunal diverticulosis, ileocolic fistula, anatomic blind loop, intestinal stricture, etc. Ileal resection and Crohn's disease Selective malabsorption with proteinuria Tropical sprue Transcobalamin II deficiency Fish tapeworm

Formerly, the pathogenesis of B_{12} malabsorption was distinguishable based on the results of a Schilling test in which a radioactive form of B_{12} was administered orally and its appearance in the urine was a sign of absorption. Radioactive B_{12} is no longer available, and Schilling tests are no longer performed. Other approaches to the differential diagnosis of B_{12} malabsorption are now employed.

Pernicious Anemia PA may be defined as a severe lack of IF due to gastric atrophy. It is a common disease in northern Europeans but occurs in all countries and ethnic groups. It is more frequent in people of African than Asian ancestry. The overall incidence is about 120 per 100,000 population in the United Kingdom (UK). The ratio of incidence in men and women among whites is ~1:1.6, and the median age of onset is 70–80 years, with only 10% of patients being <40 years of age. However, in some ethnic groups, notably blacks and Latin Americans, the age at onset of PA is generally lower. The disease occurs more commonly than by chance in close relatives and in persons with other organ-specific autoimmune diseases, for example, thyroid diseases, vitiligo, hypoparathyroidism, type 1 diabetes, and Addison's disease. It is also associated with hypogammaglobulinemia, premature graying or blue eyes, and persons of blood group A. An association with human leukocyte antigen (HLA) 3 has been reported in some but not all series and, in those with endocrine disease, with HLA-B8, -B₁₂, and -BW15. Life expectancy is normal in women once regular treatment has begun. Men had a slightly subnormal life expectancy as a result of a higher incidence of carcinoma of the stomach than in control subjects, but

current data on their life expectancy are unavailable. Gastric output of hydrochloric acid, pepsin, and IF is severely reduced. The serum gastrin level is raised, and serum pepsinogen I levels are low.

Gastric Biopsy A single endoscopic examination is recommended if PA is diagnosed. Gastric biopsy usually shows atrophy of all layers of the body and fundus, with loss of glandular elements, an absence of parietal and chief cells and replacement by mucous cells, a mixed inflammatory cell infiltrate, and perhaps intestinal metaplasia. The infiltrate of plasma cells and lymphocytes contains an excess of CD4 cells. These are directed against gastric H/K-ATPase. The antral mucosa is usually well preserved. *Helicobacter pylori* infection occurs infrequently in PA, but it has been suggested that *H. pylori* gastritis occurs at an early phase of atrophic gastritis and presents in younger patients as iron-deficiency anemia but in older patients as PA. *H. pylori* is suggested to stimulate an autoimmune process directed against parietal cells, with the *H. pylori* infection then being gradually replaced, in some individuals, by an autoimmune process.

Serum Antibodies Two types of IF immunoglobulin G antibody may be found in the sera of patients with PA. The “blocking,” or type I, antibody prevents the combination of IF and cobalamin, whereas the “binding,” or type II, antibody prevents attachment of IF to ileal mucosa. Type I occurs in the sera of ~55% of patients, and type II in 35%. IF antibodies cross the placenta and may cause temporary IF deficiency in a newborn infant. Patients with PA also show cell-mediated immunity to IF. Type I antibody has been detected rarely in the sera of patients without PA but with thyrotoxicosis, myxedema, Hashimoto's disease, or diabetes mellitus and in relatives of PA patients. IF antibodies also have been detected in gastric juice in ~80% of PA patients. These gastric antibodies may reduce absorption of dietary cobalamin by combining with small amounts of remaining IF.

Parietal cell antibody is present in the sera of almost 90% of adult patients with PA but is frequently present in other subjects. Thus, it occurs in as many as 16% of randomly selected female subjects age >60 years. The parietal cell antibody is directed against the α and β subunits of the gastric proton pump (H^+, K^+ -ATPase).

JUVENILE PERNICIOUS ANEMIA

This usually occurs in older children and resembles PA of adults. Gastric atrophy, achlorhydria, and serum IF antibodies are all present, although parietal cell antibodies are usually absent. About one-half of these patients show an associated endocrinopathy such as autoimmune thyroiditis, Addison's disease, or hypoparathyroidism; in some, mucocutaneous candidiasis occurs.

CONGENITAL INTRINSIC FACTOR DEFICIENCY OR FUNCTIONAL ABNORMALITY

An affected child usually presents with megaloblastic anemia in the first to third year of life; a few have presented as late as the second decade. The child usually has no demonstrable IF but has a normal gastric mucosa and normal secretion of acid. The inheritance is autosomal recessive. Parietal cell and IF antibodies are absent. Variants have been described in which the child is born with IF that can be detected immunologically but is unstable or functionally inactive, unable to bind cobalamin or to facilitate its uptake by ileal receptors.

GASTRECTOMY

After total gastrectomy, cobalamin deficiency is inevitable, and prophylactic cobalamin therapy should be commenced immediately after the operation. After partial gastrectomy, 10–15% of patients also develop this deficiency. The exact incidence and time of onset are most influenced by the size of the resection and the preexisting size of cobalamin body stores.

FOOD COBALAMIN MALABSORPTION

Failure of release of cobalamin from binding proteins in food is believed to be responsible for this condition, which is more common in the elderly. It is associated with low serum cobalamin levels, with or without raised serum levels of MMA and homocysteine. Typically,

TABLE 99-4 Malabsorption of Cobalamin May Occur in the Following Conditions but Is Not Usually Sufficiently Severe and Prolonged to Cause Megaloblastic Anemia

Gastric causes
Simple atrophic gastritis (food cobalamin malabsorption)
Zollinger-Ellison syndrome
Gastric bypass or bariatric surgery
Use of proton pump inhibitors
Intestinal causes
Gluten-induced enteropathy
Severe pancreatitis
HIV infection
Radiotherapy
Graft-versus-host disease
Deficiencies of cobalamin, folate, protein, ?riboflavin, ?nicotinic acid
Therapy with colchicine, para-aminosalicylate, neomycin, slow-release potassium chloride, anticonvulsant drugs, metformin, ^a cytotoxic drugs
Alcohol

^aIt is now thought that metformin lowers serum vitamin B_{12} level by lowering the level of transcobalamin I.

these patients have normal cobalamin absorption, as measured with crystalline cobalamin, but show malabsorption when a modified test using food-bound cobalamin is used. It is usually due to mild forms of atrophic gastritis or therapy with proton pump inhibitors. Bariatric surgery is likely to be an increasing cause of this form of B_{12} malabsorption and deficiency. The frequency of progression to severe cobalamin deficiency and the reasons for this progression are not clear.

■ INTESTINAL CAUSES OF COBALAMIN MALABSORPTION

Intestinal Stagnant Loop Syndrome Malabsorption of cobalamin occurs in a variety of intestinal lesions in which there is colonization of the upper small intestine by fecal organisms. This may occur in patients with jejunal diverticulosis, enteroanastomosis, or an intestinal stricture or fistula or with an anatomic blind loop due to Crohn's disease, tuberculosis, or an operative procedure.

Ileal Resection Removal of ≥ 1.2 m of terminal ileum causes malabsorption of cobalamin. In some patients after ileal resection, particularly if the ileocecal valve is incompetent, colonic bacteria may contribute further to the onset of cobalamin deficiency.

Selective Malabsorption of Cobalamin with Proteinuria (Imerslund's Syndrome; Imerslund-Gräsbeck Syndrome; Congenital Cobalamin Malabsorption; Autosomal Recessive Megaloblastic Anemia; MGA1) This autosomal recessive disease is the most common cause of megaloblastic anemia due to cobalamin deficiency in infancy in Western countries. More than 200 cases have been reported with familial clusters in Finland, Norway, the Middle East, and North Africa. The patients secrete normal amounts of IF and gastric acid but are unable to absorb cobalamin. In Finland, impaired synthesis, processing, or ligand binding of cubilin due to inherited mutations is found. In Norway, mutation of the gene for AMN has been reported. Other tests of intestinal absorption are normal. Over 90% of these patients show nonspecific proteinuria, but renal function is otherwise normal, and renal biopsy has not shown any consistent renal defect. A few have shown aminoaciduria and congenital renal abnormalities, such as duplication of the renal pelvis.

Tropical Sprue Nearly all patients with acute and subacute tropical sprue show malabsorption of cobalamin; this may persist as the principal abnormality in the chronic form of the disease, when the patient may present with megaloblastic anemia or neuropathy due to cobalamin deficiency. Absorption of cobalamin usually improves after antibiotic therapy and, in the early stages, folic acid therapy.

Fish Tapeworm Infestation The fish tapeworm (*Diphyllobothrium latum*) lives in the small intestine of humans and accumulates cobalamin from food, rendering the cobalamin unavailable for absorption. Individuals acquire the worm by eating raw or partly cooked fish. Infestation is common around the lakes of Scandinavia, Germany, Japan, North America, and Russia. Megaloblastic anemia or cobalamin neuropathy occurs only in those with a heavy infestation.

Gluten-Induced Enteropathy Malabsorption of cobalamin occurs in ~30% of untreated patients (presumably those in whom the disease extends to the ileum). Cobalamin deficiency is not severe in these patients and is corrected with a gluten-free diet.

Severe Chronic Pancreatitis In this condition, lack of trypsin is thought to cause dietary cobalamin attached to gastric non-IF (R) binder to be unavailable for absorption. It also has been proposed that in pancreatitis, the concentration of calcium ions in the ileum falls below the level needed to maintain normal cobalamin absorption.

HIV Infection Serum cobalamin levels tend to fall in patients with HIV infection and are subnormal in 10–35% of those with AIDS. Malabsorption of cobalamin not corrected by IF has been shown in some, but not all, patients with subnormal serum cobalamin levels. Cobalamin deficiency sufficiently severe to cause megaloblastic anemia or neuropathy is rare.

Zollinger-Ellison Syndrome Malabsorption of cobalamin has been reported in the Zollinger-Ellison syndrome. It is thought that there is a failure to release cobalamin from R-binding protein due to inactivation of pancreatic trypsin by high acidity, as well as interference with IF binding of cobalamin.

Radiotherapy Both total-body irradiation and local radiotherapy to the ileum (e.g., as a complication of radiotherapy for carcinoma of the cervix) may cause malabsorption of cobalamin.

Graft-versus-Host Disease This commonly affects the small intestine. Malabsorption of cobalamin due to abnormal gut flora, as well as damage to ileal mucosa, is common.

Drugs The drugs that have been reported to cause malabsorption of cobalamin are listed in Table 99-4. However, megaloblastic anemia due to these drugs is rare. It has been suggested that metformin lowers serum B_{12} by lowering TC I level rather than causing malabsorption of B_{12} .

■ ABNORMALITIES OF COBALAMIN METABOLISM

Congenital Transcobalamin II Deficiency or Abnormality Infants with TC II deficiency usually present with megaloblastic anemia within a few weeks of birth. Serum cobalamin and folate levels are normal, but the anemia responds to massive (e.g., 1 mg three times weekly) injections of cobalamin. Some cases show neurologic complications. The protein may be present but functionally inert. Genetic abnormalities found include mutations of an intraexonic cryptic splice site, extensive deletion, single nucleotide deletion, nonsense mutation, and an RNA editing defect. Malabsorption of cobalamin occurs in all cases, and serum immunoglobulins are usually reduced. Failure to institute adequate cobalamin therapy or treatment with folic acid may lead to neurologic damage.

Congenital Methylmalonic Acidemia and Aciduria Infants with this abnormality are ill from birth with vomiting, failure to thrive, severe metabolic acidosis, ketosis, and mental retardation. Anemia, if present, is normocytic and normoblastic. The condition may be due to a functional defect in either mitochondrial methylmalonyl-CoA mutase or its cofactor adocobalamin. Mutations in the methylmalonyl-CoA mutase are not responsive or are only poorly responsive to treatment with cobalamin. A proportion of infants with failure of adocobalamin synthesis respond to cobalamin in large doses. Some children have combined methylmalonic aciduria and homocystinuria due to defective formation of both cobalamin coenzymes. This usually presents in the first year of life with feeding difficulties, developmental delay, microcephaly, seizures, hypotonia, and megaloblastic anemia.

Acquired Abnormality of Cobalamin Metabolism: Nitrous Oxide Inhalation Nitrous oxide (N_2O) irreversibly oxidizes methylcobalamin to an inactive precursor; this inactivates methionine synthase. Megaloblastic anemia has occurred in patients undergoing prolonged N_2O anesthesia (e.g., in intensive care units). A neuropathy resembling cobalamin neuropathy has been described in dentists and anesthetists who are exposed repeatedly to N_2O . Methylmalonic aciduria does not occur as adocobalamin is not inactivated by N_2O .

CAUSES OF FOLATE DEFICIENCY

(Table 99-5)

■ NUTRITIONAL

Dietary folate deficiency is common. Indeed, in most patients with folate deficiency, a nutritional element is present. Certain individuals are particularly prone to have diets containing inadequate amounts of folate (Table 99-5). In the United States and other countries where fortification of the diet with folic acid has been adopted, the prevalence of folate deficiency has dropped dramatically and is now almost restricted to high-risk groups with increased folate needs. Nutritional folate deficiency occurs in kwashiorkor and scurvy and in infants with repeated infections or those who are fed solely on goats' milk, which has a low folate content.

TABLE 99-5 Causes of Folate Deficiency**Dietary^a**

Particularly in: old age, infancy, poverty, alcoholism, chronic invalids, and the psychiatrically disturbed; may be associated with scurvy or kwashiorkor

Malabsorption**Major causes of deficiency**

Tropical sprue, gluten-induced enteropathy in children and adults, and in association with dermatitis herpetiformis, specific malabsorption of folate, intestinal megaloblastosis caused by severe cobalamin or folate deficiency

Minor causes of deficiency

Extensive jejunal resection, Crohn's disease, partial gastrectomy, congestive heart failure, Whipple's disease, scleroderma, amyloid, diabetic enteropathy, systemic bacterial infection, lymphoma, sulfasalazine (Salazopyrin)

Excess utilization or loss**Physiologic**

Pregnancy and lactation, prematurity

Pathologic

Hematologic diseases: chronic hemolytic anemias, sickle cell anemia, thalassemia major, myelofibrosis

Malignant diseases: carcinoma, lymphoma, leukemia, myeloma

Inflammatory diseases: tuberculosis, Crohn's disease, psoriasis, exfoliative dermatitis, malaria

Metabolic disease: homocystinuria

Excess urinary loss: congestive heart failure, active liver disease

Hemodialysis, peritoneal dialysis

Antifolate drugs^b

Anticonvulsant drugs (phenytoin, primidone, barbiturates), sulfasalazine

Nitrofurantoin, tetracycline, antituberculosis (less well documented)

Mixed causes

Liver diseases, alcoholism, intensive care units

^aIn severely folate-deficient patients with causes other than those listed under Dietary, poor dietary intake is often present. ^bDrugs inhibiting dihydrofolate reductase are discussed in the text.

MALABSORPTION

Malabsorption of dietary folate occurs in tropical sprue and in gluten-induced enteropathy. In the rare congenital recessive syndrome of selective malabsorption of folate due to mutation of the PCFT, there is an associated defect of folate transport into the cerebrospinal fluid, and these patients show megaloblastic anemia, which responds to physiologic doses of folic acid given parenterally but not orally. They also show mental retardation, convulsions, and other central nervous system abnormalities. Minor degrees of malabsorption may also occur after jejunal resection or partial gastrectomy, in Crohn's disease, and in systemic infections, but in these conditions, if severe deficiency occurs, it is usually largely due to poor nutrition. Malabsorption of folate has been described in patients receiving sulfasalazine (Salazopyrin), cholestyramine, and triamterene.

EXCESS UTILIZATION OR LOSS

Pregnancy Folate requirements are increased by 200–300 µg to ~400 µg daily in a normal pregnancy, partly because of transfer of the vitamin to the fetus but mainly because of increased folate catabolism due to cleavage of folate coenzymes in rapidly proliferating tissues. Megaloblastic anemia due to this deficiency is prevented by prophylactic folic acid therapy. It occurred in 0.5% of pregnancies in the UK and other Western countries before prophylaxis with folic acid, but the incidence is much higher in countries where the general nutritional status is poor.

Prematurity A newborn infant, whether full term or premature, has higher serum and red cell folate concentrations than does an adult. However, a newborn infant's demand for folate has been estimated to

be up to 10 times that of adults on a weight basis, and the neonatal folate level falls rapidly to the lowest values at about 6 weeks of age. The falls are steepest and are liable to reach subnormal levels in premature babies, a number of whom develop megaloblastic anemia responsive to folic acid at about 4–6 weeks of age. This occurs particularly in the smallest babies (<1500 g birth weight) and those who have feeding difficulties or infections or have undergone multiple exchange transfusions. In these babies, prophylactic folic acid should be given.

Hematologic Disorders Folate deficiency frequently occurs in chronic hemolytic anemia, particularly in sickle cell disease, autoimmune hemolytic anemia, and congenital spherocytosis. In these and other conditions of increased cell turnover (e.g., myelofibrosis, malignancies), folate deficiency arises because it is not completely reutilized after performing coenzyme functions.

Inflammatory Conditions Chronic inflammatory diseases such as tuberculosis, rheumatoid arthritis, Crohn's disease, psoriasis, exfoliative dermatitis, bacterial endocarditis, and chronic bacterial infections cause deficiency by reducing the appetite and increasing the demand for folate. Systemic infections also may cause malabsorption of folate. Severe deficiency is virtually confined to the patients with the most active disease and the poorest diet.

Homocystinuria This is a rare metabolic defect in the conversion of homocysteine to cystathione. Folate deficiency occurring in most of these patients may be due to excessive utilization because of compensatory increased conversion of homocysteine to methionine.

Long-Term Dialysis Because folate is only loosely bound to plasma proteins, it is easily removed from plasma by dialysis. In patients with anorexia, vomiting, infections, and hemolysis, folate stores are particularly likely to become depleted. Routine folate prophylaxis is now given.

Congestive Heart Failure and Liver Disease Excess urinary folate losses of >100 µg per day may occur in some of these patients. The explanation appears to be release of folate from damaged liver cells.

■ ANTIFOLATE DRUGS

A large number of epileptics who are receiving long-term therapy with phenytoin or primidone, with or without barbiturates, develop low serum and red cell folate levels. The exact mechanism is unclear. Alcohol may also be a folate antagonist, as patients who are drinking spirits may develop megaloblastic anemia that will respond to normal quantities of dietary folate or to physiologic doses of folic acid only if alcohol is withdrawn. Macrocytosis of red cells is associated with chronic alcohol intake even when folate levels are normal. Inadequate folate intake is the major factor in the development of deficiency in spirit-drinking alcoholics. Beer is relatively folate-rich in some countries, depending on the technique used for brewing.

The drugs that inhibit DHF reductase include methotrexate, pyrimethamine, and trimethoprim. Methotrexate has the most powerful action against the human enzyme, whereas trimethoprim is most active against the bacterial enzyme and is likely to cause megaloblastic anemia only when used in conjunction with sulfamethoxazole in patients with preexisting folate or cobalamin deficiency. The activity of pyrimethamine is intermediate. The antidote to these drugs is folinic acid (5-formyl-THF).

■ CONGENITAL ABNORMALITIES OF FOLATE METABOLISM

Some infants with congenital defects of folate enzymes (e.g., cyclohydrolase or methionine synthase) have had megaloblastic anemia.

DIAGNOSIS OF COBALAMIN AND FOLATE DEFICIENCIES

The diagnosis of cobalamin or folate deficiency has traditionally depended on the recognition of the relevant abnormalities in the peripheral blood and analysis of the blood levels of the vitamins.

■ COBALAMIN DEFICIENCY

Serum Cobalamin This is measured by an automated enzyme-linked immunosorbent assay (ELISA) or competitive-binding luminescence assay (CBLA). Normal serum levels range from 118–148 pmol/L (160–200 ng/L) to ~738 pmol/L (1000 ng/L). In patients with megaloblastic anemia due to cobalamin deficiency, the level is usually <74 pmol/L (100 ng/L). In general, the more severe the deficiency, the lower is the serum cobalamin level. In patients with spinal cord damage due to the deficiency, levels are very low even in the absence of anemia. Values between 74 and 148 pmol/L (100 and 200 ng/L) are regarded as borderline. They may occur, for instance, in pregnancy, in patients with megaloblastic anemia due to folate deficiency. They may also be due to heterozygous, homozygous, or compound heterozygous mutations of the gene *TCN1* that codes for HC (TC I). There is then no clinical or hematologic abnormality. The serum cobalamin level is sufficiently robust, cost-effective, and most convenient to rule out cobalamin deficiency in the vast majority of patients suspected of having this problem. However, problems have arisen with commercial CBLA assays involving IF in PA patients with intrinsic antibodies in serum. These antibodies may cause false normal serum vitamin B₁₂ levels in up to 50% of cases tested. Where clinical indications of PA are strong, a normal serum vitamin B₁₂ does not rule out the diagnosis. Serum MMA levels will be elevated in untreated PA (see below).

Folate deficiency, TC I (HC) deficiency, oral contraceptives, and multiple myeloma have all been associated with low serum B₁₂ levels that do not indicate B₁₂ deficiency. On the other hand, high serum B₁₂ levels are usually due to raised serum TC I levels and can be due to the presence of liver, renal, or myeloproliferative diseases or to cancer of the breast, colon, or liver.

Serum Methylmalonate and Homocysteine In patients with cobalamin deficiency sufficient to cause anemia or neuropathy, the serum MMA level is raised. Sensitive methods for measuring MMA and homocysteine in serum have been introduced and recommended for the early diagnosis of cobalamin deficiency, even in the absence of hematologic abnormalities or subnormal levels of serum cobalamin. Serum MMA levels fluctuate, however, in patients with renal failure. Mildly elevated serum MMA and/or homocysteine levels occur in up to 30% of apparently healthy volunteers, with serum cobalamin levels up to 258 pmol/L (350 ng/L) and normal serum folate levels; 15% of elderly subjects, even with cobalamin levels >258 pmol/L (>350 ng/L), have this pattern of raised metabolite levels. These findings bring into question the exact cutoff points for normal MMA and homocysteine levels. It is also unclear at present whether these mildly raised metabolite levels have clinical consequences.

Serum homocysteine is raised in both early cobalamin and folate deficiency but may be raised in other conditions, for example, chronic renal disease, alcoholism, smoking, pyridoxine deficiency, hypothyroidism, and therapy with steroids, cyclosporine, and other drugs. Levels are also higher in serum than in plasma, in men than in premenopausal women, in women taking hormone replacement therapy or in oral contraceptive users, and in elderly persons and patients with several inborn errors of metabolism affecting enzymes in trans-sulfuration pathways of homocysteine metabolism. Thus, homocysteine levels must be carefully interpreted for diagnosis of cobalamin or folate deficiency.

Tests for the Cause of Cobalamin Deficiency Only vegans, strict vegetarians, or people living on a totally inadequate diet will become vitamin B₁₂ deficient because of inadequate intake. Studies of cobalamin absorption once were widely used, but difficulty in obtaining radioactive cobalamin and ensuring that IF preparations are free of viruses has made these tests obsolete. Tests to diagnose PA include serum gastrin, which is raised; serum pepsinogen I, which is low in PA (90–92%) but also in other conditions; and gastric endoscopy. Tests for IF and parietal cell antibodies are also used, as well as tests for individual intestinal diseases.

Patients with atrophic gastritis may also have sufficient occult gastrointestinal blood loss to have iron deficiency as well as vitamin B₁₂

deficiency. Iron deficiency may blunt the development of macrocytosis when iron deficiency and B₁₂ deficiency coexist. Iron deficiency is much more common than B₁₂ deficiency, and in people older than age 60 years, B₁₂ deficiency may accompany iron deficiency in 15–20% of cases. Thus, patients diagnosed with iron-deficiency anemia should have B₁₂ levels assessed, and those diagnosed with B₁₂ deficiency should have their iron status assessed.

■ FOLATE DEFICIENCY

Serum Folate This is also measured by an ELISA technique. In most laboratories, the normal range is from 11 nmol/L (2 µg/L) to ~82 nmol/L (15 µg/L). The serum folate level is low in all folate-deficient patients. It also reflects recent diet. Because of this, serum folate may be low before there is hematologic or biochemical evidence of deficiency. Serum folate rises in severe cobalamin deficiency because of the block in conversion of MTHF to THF inside cells; raised levels have also been reported in the intestinal stagnant loop syndrome due to absorption of bacterially synthesized folate.

Red Cell Folate The red cell folate assay is a valuable test of body folate stores. It is less affected than the serum assay by recent diet and traces of hemolysis. In normal adults, concentrations range from 880 to 3520 µmol/L (160–640 µg/L) of packed red cells. Subnormal levels occur in patients with megaloblastic anemia due to folate deficiency but also in nearly two-thirds of patients with severe cobalamin deficiency. False-normal results may occur if a folate-deficient patient has received a recent blood transfusion or if a patient has a raised reticulocyte count. Serum homocysteine assay is discussed earlier.

Tests for the Cause of Folate Deficiency The diet history is important. Tests for transglutaminase antibodies are performed to confirm or exclude celiac disease. If positive, duodenal biopsy is needed. An underlying disease causing increased folate breakdown should also be excluded.

TREATMENT

Cobalamin and Folate Deficiency

It is usually possible to establish which of the two deficiencies, folate or cobalamin, is the cause of the anemia and to treat only with the appropriate vitamin. In patients who enter the hospital severely ill, however, it may be necessary to treat with both vitamins in large doses once blood samples have been taken for cobalamin and folate assays and a bone marrow biopsy has been performed (if deemed necessary). Transfusion is usually unnecessary and inadvisable. If it is essential, packed red cells should be given slowly, one or two units only, with the usual treatment for heart failure if present. Potassium supplements have been recommended to obviate the danger of the hypokalemia but are not necessary. Occasionally, an excessive rise in platelets occurs after 1–2 weeks of therapy. Antiplatelet therapy, for example, aspirin, should be considered if the platelet count rises to >800 × 10⁹/L.

COBALAMIN DEFICIENCY

It is usually necessary to treat patients who have developed cobalamin deficiency with lifelong regular cobalamin injections. In the UK, the form used is hydroxocobalamin; in the United States, cyanocobalamin. In a few instances, the underlying cause of cobalamin deficiency can be permanently corrected, for example, fish tapeworm, tropical sprue, or an intestinal stagnant loop that is amenable to surgery. The indications for starting cobalamin therapy are a well-documented megaloblastic anemia or other hematologic abnormalities and neuropathy due to the deficiency. Patients with borderline serum cobalamin levels but no hematologic or other abnormality may be followed to make sure that the cobalamin deficiency does not progress (see below). If malabsorption of cobalamin or rises in serum MMA levels have been demonstrated, however, these patients also should be given regular maintenance cobalamin

therapy. Cobalamin should be given routinely to all patients who have had a total gastrectomy or ileal resection. Patients who have undergone gastric reduction for control of obesity or who are receiving long-term treatment with proton pump inhibitors should be screened and, if necessary, given cobalamin replacement.

Replenishment of body stores should be complete with six 1000- μ g IM injections of hydroxocobalamin given at 3- to 7-day intervals. More frequent doses are usually used in patients with cobalamin neuropathy, but there is no evidence that they produce a better response. Allergic reactions are rare and may require desensitization or antihistamine or glucocorticoid cover. For maintenance therapy, 1000 μ g hydroxocobalamin IM once every 3 months is satisfactory. Because of the poorer retention of cyanocobalamin, protocols generally use higher and more frequent doses, for example, 1000 μ g IM, monthly, for maintenance treatment.

Because a small fraction of cobalamin can be absorbed passively through mucous membranes even when there is complete failure of physiologic IF-dependent absorption, large daily oral doses (1000–2000 μ g) of cyanocobalamin are used in PA for replacement (especially in Canada and Sweden) and maintenance of normal cobalamin status in, for example, food malabsorption of cobalamin. Sublingual therapy has also been proposed for those in whom injections are difficult because of a bleeding tendency and who may not tolerate oral therapy. If oral therapy is used, it is important to monitor compliance, particularly with elderly, forgetful patients. This author prefers parenteral therapy for initial treatment, particularly in severe anemia or if a neuropathy is present, and for maintenance in PA. Oral B_{12} therapy even with low doses of 50 μ g daily may have a larger role in treating food malabsorption of B_{12} .

For treatment of patients with subnormal serum vitamin B_{12} levels with a normal MCV and no hypersegmentation of neutrophils, a negative IF antibody test in the absence of tests of B_{12} absorption is problematic. Some (perhaps 15%) cases may be due to TC I (HC) deficiency. Homocysteine and/or MMA measurements may help, but in the absence of these tests and with otherwise normal gastrointestinal function, repeat serum B_{12} assay after 6–12 months may help one decide whether to start cobalamin therapy.

Vitamin B_{12} injections are used in a wide variety of diseases, often neurologic, despite normal serum B_{12} and folate levels and a normal blood count and in the absence of randomized, double-blind, controlled trials. These conditions include multiple sclerosis and chronic fatigue syndrome/myalgic encephalomyelitis (ME). It seems probable that any benefit is due to the placebo effect of a usually painless, pink injection. In ME, oral B_{12} therapy, despite providing equally large amounts of B_{12} , has not been beneficial, supporting the view of the effect of the injections being placebo only.

FOLATE DEFICIENCY

Oral doses of 5–15 mg of folic acid daily are satisfactory, as sufficient folate is absorbed from these extremely large doses even in patients with severe malabsorption. The length of time therapy must be continued depends on the underlying disease. It is customary to continue therapy for about 4 months, when all folate-deficient red cells will have been eliminated and replaced by new folate-replete populations.

Before large doses of folic acid are given, cobalamin deficiency must be excluded and, if present, corrected; otherwise, cobalamin neuropathy may develop despite a response of the anemia of cobalamin deficiency to folate therapy. Studies in the United States, however, suggest that there is no increase in the proportion of individuals with low serum cobalamin levels and no anemia since food fortification with folic acid, but it is unknown if there has been a change in incidence of cobalamin neuropathy.

Long-term folic acid therapy is required when the underlying cause of the deficiency cannot be corrected and the deficiency is likely to recur, for example, in chronic dialysis or hemolytic anemias. It may also be necessary in gluten-induced enteropathy that does not respond to a gluten-free diet. Where mild but chronic folate deficiency occurs, it is preferable to encourage improvement

in the diet after correcting the deficiency with a short course of folic acid. In any patient receiving long-term folic acid therapy, it is important to measure the serum cobalamin level at regular (e.g., once-yearly) intervals to exclude the coincidental development of cobalamin deficiency.

Folinic Acid (5-Formyl-THF) This is a stable form of fully reduced folate. It is given orally or parenterally to overcome the toxic effects of methotrexate or other DHF reductase inhibitors, for example, trimethoprim or cotrimoxazole.

PROPHYLACTIC FOLIC ACID

Prophylactic folic acid is used in chronic dialysis patients and in parenteral feeds. Prophylactic folic acid has been used to reduce homocysteine levels to prevent cardiovascular disease and for cognitive function in the elderly, but there are no firm data to show any benefit.

Pregnancy In over 70 countries (but none in Europe), food is fortified with folic acid (in grain or flour) to reduce the risk of NTDs. Nevertheless, folic acid, 400 μ g daily, should be given as a supplement before and throughout pregnancy to prevent megaloblastic anemia and reduce the incidence of NTDs, even in countries with fortification of the diet. The levels of fortification provide up to 400 μ g daily on average in Chile, but in most countries, it is nearer to 200 μ g, so periconceptual folic acid is still needed. Most if not all the folic acid used in fortification and eaten over three meals a day will be converted during absorption to methyltetrahydrofolate. This compound will not correct the anemia in B_{12} deficiency. Studies in early pregnancy show significant lack of compliance with the folic acid supplements, emphasizing the benefit of food fortification. Supplemental folic acid reduces the incidence of birth defects in babies born to diabetic mothers. In women who have had a previous fetus with an NTD, a dose of 5 mg daily is recommended when pregnancy is contemplated and throughout the subsequent pregnancy.

Infancy and Childhood The incidence of folate deficiency is so high in the smallest premature babies during the first 6 weeks of life that folic acid (e.g., 1 mg daily) should be given routinely to those weighing <1500 g at birth and to larger premature babies who require exchange transfusions or develop feeding difficulties, infections, or vomiting and diarrhea.

The World Health Organization currently recommends routine supplementation with iron and folic acid in children in countries where iron deficiency is common and child mortality, largely due to infectious diseases, is high. However, some studies suggest that in areas where malaria rates are high, this approach may increase the incidence of severe illness and death. Even where malaria is rare, there appears to be no survival benefit.

MEGALOBLASTIC ANEMIA NOT DUE TO COBALAMIN OR FOLATE DEFICIENCY OR ALTERED METABOLISM

This may occur with many antimetabolic drugs (e.g., hydroxyurea, cytosine arabinoside, 6-mercaptopurine) that inhibit DNA replication. Antiviral nucleoside analogues used in treatment of HIV infection may also cause macrocytosis and megaloblastic marrow changes. In the rare disease orotic aciduria, two consecutive enzymes in purine synthesis are defective. The condition responds to therapy with uridine, which bypasses the block. In thiamine-responsive megaloblastic anemia, there is a genetic defect in the high-affinity thiamine transport (*SLC19A2*) gene. This causes defective RNA ribose synthesis through impaired activity of transketolase, a thiamine-dependent enzyme in the pentose cycle. This leads to reduced nucleic acid production. It may be associated with diabetes mellitus and deafness and the presence of many ringed sideroblasts in the marrow. The explanation is unclear for megaloblastic changes in the marrow in some patients with acute myeloid leukemia and myelodysplasia.

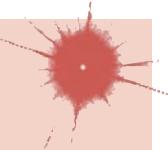
FURTHER READING

- BERRY RJ: Lack of historical evidence to support folic acid exacerbation of the neuropathy caused by vitamin B12 deficiency. *Am J Clin Nutr* 110:554, 2019.
- BUNN HF: Vitamin B12 and pernicious anemia: The dawn of molecular medicine. *N Engl J Med* 370:773, 2014.
- DEL BO C et al: Effect of two different sublingual dosages of vitamin B12 on cobalamin nutritional status in vegans and vegetarians with a marginal deficiency: A randomized controlled trial. *Clin Nutr* 38:575, 2019.
- GREEN R: Vitamin B12 deficiency from the perspective of a practicing hematologist. *Blood* 129:2603, 2017.
- GREEN R et al: Vitamin B12 deficiency. *Nat Rev Dis Primers* 3:17040, 2017.
- HESDORFFER CS, LONGO DL: Drug-induced megaloblastic anemia. *N Engl J Med* 373:1649, 2015.
- MA F et al: Effects of folic acid and vitamin B12 alone and in combination on cognitive function and inflammatory factors in the elderly with mild cognitive impairment: A single blind experimental design. *Curr Alzheimer Res* 16: 622, 2019.
- MILLER JW: Proton pump inhibitors, H2-receptor antagonists, metformin, and vitamin B-12 deficiency: Clinical implications. *Adv Nutr* 9:511S, 2018.
- O'CONNOR DMA et al: Plasma concentrations of vitamin B12 (B12) and folate and global cognitive function in an older population: cross-sectional findings from the Irish Longitudinal Study on Ageing (TILDA). *Br J Nutr* 124:602, 2020.
- ROGERS LM et al: Global folate status in women of reproductive age: A systematic review with emphasis on methodological issues. *Ann N Y Acad Sci* 1431:35, 2018.
- SALINAS M et al: High frequency of anti-parietal cell antibody (APCA) and intrinsic factor blocking antibody (IFBA) in individuals with severe vitamin B12 deficiency: An observational study in primary care patients. *Clin Chem Lab Med* 58:424, 2020.
- ZARIC BL et al: Homocysteine and hyperhomocysteinaemia. *Curr Med Chem* 26:2948, 2019.

100

Hemolytic Anemias

Lucio Luzzatto, Lucia De Franceschi



DEFINITIONS

A finite life span is a distinct characteristic of red cells. Hence, a logical, time-honored classification of anemias is in three groups: (1) decreased production of red cells, (2) increased destruction of red cells, and (3) acute blood loss. Decreased production is covered in [Chaps. 97, 98, and 102](#); acute blood loss in [Chap. 101](#); increased destruction is covered in this chapter.

All patients who are anemic as a result of either increased destruction of red cells or acute blood loss have one important element in common: the anemia results from overconsumption of red cells from the peripheral blood, whereas the supply of cells from the bone marrow is normal (indeed, it is usually increased). However, with blood loss, as in acute hemorrhage, the red cells are physically lost *from* the body itself; this is fundamentally different from destruction of red cells *within* the body, as in hemolytic anemias (HAs).

With respect to primary etiology, HAs may be *inherited* or *acquired*; from a clinical point of view, they may be more *acute* or more *chronic*, and they may vary from mild to very severe; the site of hemolysis may be predominantly *intravascular* or *extravascular*. With respect to mechanisms, HAs may be due to *intracorporeal* causes or to *extracorporeal* causes ([Table 100-1](#)). But before reviewing the individual types

TABLE 100-1 Classification of Hemolytic Anemias^a

	INTRACORPOERAL DEFECTS	EXTRACORPOERAL FACTORS
Inherited	Hemoglobinopathies Enzymopathies Membrane-cytoskeletal defects	Familial (atypical) hemolytic-uremic syndrome
Acquired	Paroxysmal nocturnal hemoglobinuria (PNH)	Mechanical destruction (microangiopathic) Toxic agents Drugs Infectious Autoimmune

^aHereditary causes correlate with intracorporeal defects because these defects are due to inherited mutations; the one exception is PNH because the defect is due to an acquired somatic mutation. Conversely, acquired causes correlate with extracorporeal factors because mostly these factors are exogenous; the one exception is familial hemolytic-uremic syndrome (HUS; often referred to as atypical HUS) because here an inherited abnormality permits complement activation triggered by exogenous factors, to become excessive, with bouts of production of membrane attack complex capable of destroying normal red cells. Interestingly, in both PNH and aHUS hemolysis is complement-mediated.

of HA, it is appropriate to consider what general features they have in common, in terms of clinical aspects and pathophysiology.

GENERAL CLINICAL AND LABORATORY FEATURES

The clinical presentation of a patient with anemia is greatly influenced in the first place by whether the onset is abrupt or gradual and HAs are no exception. A patient with autoimmune HA or with favism may be a medical emergency, whereas a patient with mild hereditary spherocytosis (HS) or with cold agglutinin disease (CAD) may be diagnosed after years. This is due in large measure to the remarkable ability of the body to adapt to anemia when it is slowly progressing ([Chap. 63](#)).

What differentiates HAs from other anemias is that the patient has signs and symptoms arising directly from hemolysis ([Table 100-2](#)). At the clinical level, the main sign is *jaundice*; in addition, the patient may report discoloration of the urine. In many cases of HA, the spleen is enlarged because it is a preferential site of hemolysis; and in some cases, the liver may be enlarged as well. In all severe congenital forms of HA, there may also be skeletal changes due to overactivity of the bone marrow: they are never as severe as in thalassemia major because there is less ineffective erythropoiesis, or none at all.

The laboratory features of HA are related to (i) hemolysis per se, and (ii) the erythropoietic response of the bone marrow. In most cases hemolysis is largely extravascular, and it produces an increase in unconjugated bilirubin and aspartate aminotransferase (AST) in the serum; urobilinogen will be increased in both urine and stool. If hemolysis is mainly intravascular, the telltale sign is hemoglobinuria (often associated with hemosiderinuria); in the serum there is free hemoglobin, lactate dehydrogenase (LDH) is increased, and haptoglobin is reduced. In contrast, the serum bilirubin level may be normal

TABLE 100-2 Features Common to Most Patients with a Hemolytic Disorder

General examination	Jaundice, pallor
Other physical findings	Spleen may be enlarged; bossing of skull in severe congenital cases
Hemoglobin level	From normal to severely reduced
MCV, MCH	Usually increased
Reticulocytes	Usually increased
Bilirubin	Almost always increased (mostly unconjugated)
LDH	Increased (up to 10× normal with intravascular hemolysis)
Haptoglobin	Reduced to absent if hemolysis is at least in part intravascular

Abbreviations: LDH, lactate dehydrogenase; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume.

or only mildly elevated. The main sign of the erythropoietic response by the bone marrow is an increase in reticulocytes (a test all too often neglected in the initial workup of a patient with anemia). Usually the increase will be reflected in both the percentage of reticulocytes (the more commonly quoted figure) and in the absolute reticulocyte count (the more definitive parameter). The increased number of reticulocytes is associated with an increased mean corpuscular volume (MCV) in the blood count. On the blood smear, this is reflected in the presence of macrocytes; there is also polychromasia, and sometimes one sees nucleated red cells. In most cases, a bone marrow aspirate is not necessary in the diagnostic workup; if it is done, it will show erythroid hyperplasia. In practice, once an HA is suspected, specific tests will usually be required for a definitive diagnosis of a specific type of HA.

■ GENERAL PATHOPHYSIOLOGY

The mature red cell is the product of a developmental pathway that brings the phenomenon of differentiation to an extreme. An orderly sequence of events produces synchronous changes, whereby the gradual accumulation of a huge amount of hemoglobin in the cytoplasm (to a final level of 340 g/L, i.e., about 5 mM) goes hand in hand with the gradual loss of cellular organelles and of biosynthetic abilities. In the end, the erythroid cell undergoes a process that has features of apoptosis, including nuclear pyknosis and eventually extrusion of the nucleus. However, the final result is more altruistic than suicidal; the cytoplasmic body, instead of disintegrating, is now able to provide oxygen to all cells in the human organism for some remaining 120 days of the red cell life span.

As a result of this unique process of differentiation and maturation, intermediary metabolism is drastically curtailed in mature red cells (**Fig. 100-1**); for instance, cytochrome-mediated oxidative phosphorylation has been lost with the loss of mitochondria (through a process of physiologic autophagy); therefore, there is no backup to anaerobic glycolysis, which in the red cell is the only provider of adenosine triphosphate (ATP). Also, the capacity of making protein has been lost with the loss of ribosomes. This places the cell's limited metabolic apparatus at risk, because if any protein component deteriorates, it cannot be replaced, as it would be in most other cells; and in fact, the activity of most enzymes gradually decreases as red cells age. At the same time, during their long time in circulation, various red cell components inevitably accumulate damage and become physically denser. The anion exchanger known as band 3 is the most abundant protein in the red cell membrane (Fig. 100-2 and Table 100-3), with about 1.2 million molecules per red cell. As red cells age and become denser, probability is increased that a region of the band 3 molecule becomes exposed on the cell surface and contributes to creating an antigenic site recognizable by low-avidity naturally occurring anti-band 3 IgG antibodies. This process might be enhanced by the clustering of band 3 molecules favored by the antibody itself and by the binding of hemichromes arising from hemoglobin degradation. Senescent red cells thus become opsonized, and this is the signal for phagocytosis by macrophages in the spleen, in the liver, and elsewhere. This process may become accelerated in various ways in HA.

Another consequence of the relative simplicity of red cells is that they have a limited range of ways to manifest distress under hardship; in essence, any sort of metabolic failure will eventually lead either to structural damage to the membrane or to failure of the cation pump. In either case, the life span of the red cell is reduced, which is the definition of a *hemolytic disorder*. If the rate of red cell destruction exceeds the capacity of the bone marrow to produce more red cells, the hemolytic disorder will manifest as HA.

Thus, the essential pathophysiologic process common to all HAs is an increased red cell turnover; in many HAs, this is due at least in part to an acceleration of the senescence process described above. The gold standard for proving that the life span of red cells is reduced (compared to the normal value of about 120 days) is a *red cell survival study*, which can be carried out by labeling the red cells with ^{51}Cr and measuring the fall in radioactivity over several days or weeks (this classic test can now be replaced by a methodology using the nonradioactive isotope ^{15}N). If the hemolytic event is transient, it does not usually cause any long-term

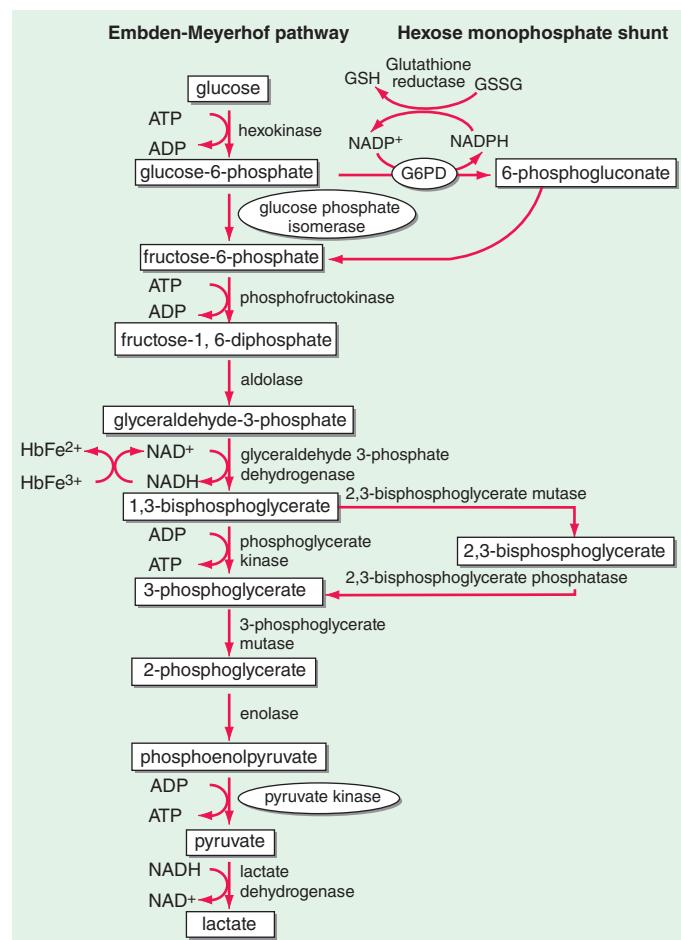


FIGURE 100-1 Red blood cell (RBC) metabolism. The Embden-Meyerhof pathway (glycolysis) generates ATP required for cation transport and for membrane maintenance. The generation of NADH maintains hemoglobin iron in a reduced state. The hexose monophosphate shunt generates NADPH that is used to reduce glutathione, which protects the red cell against oxidant stress; the 6-phosphogluconate, after decarboxylation, can be recycled via pentose sugars to glycolysis. Regulation of the 2,3-bisphosphoglycerate level is a critical determinant of oxygen affinity of hemoglobin. Enzyme deficiency states in order of prevalence: glucose-6-phosphate dehydrogenase (G6PD) > pyruvate kinase > glucose-6-phosphate isomerase > rare deficiencies of other enzymes in the pathway. The more common enzyme deficiencies are circled.

consequences, except for an increased requirement for erythropoietic factors, particularly folic acid. However, if hemolysis is recurrent or persistent, the increased bilirubin production favors the formation of gallstones. If a considerable proportion of hemolysis takes place in the spleen, as is often the case, splenomegaly may become increasingly a feature, and hypersplenism may develop, with consequent neutropenia and/or thrombocytopenia.

The increased red cell turnover has important consequences. In normal subjects, the iron from effete red cells is very efficiently recycled by the body; however, with chronic intravascular hemolysis, the persistent hemoglobinuria will cause considerable iron loss, needing replacement. With chronic extravascular hemolysis, the opposite problem, iron overload, is more common, especially if the patient needs frequent blood transfusions. Even without blood transfusion, when erythropoiesis is massively increased, the release of erythoferrone from erythroid cells suppresses hepcidin, causing increased iron absorption. In the long run, in the absence of iron-chelation therapy, iron overload will cause secondary hemochromatosis; this will cause damage particularly to the liver, eventually leading to cirrhosis; and to the heart muscle, eventually causing heart failure.

Compensated Hemolysis versus Hemolytic Anemia Red cell destruction is a potent stimulus for erythropoiesis, which is mediated

by erythropoietin (EPO) produced by the kidney. This mechanism is so effective that in many cases the increased output of red cells from the bone marrow can fully balance an increased destruction of red cells. In such cases, we say that hemolysis is *compensated*. The pathophysiology of compensated hemolysis is similar to what we have just described, except there is no anemia. This notion is important from the diagnostic point of view, because a patient with a hemolytic condition, even an inherited one, may present without anemia; and it is also important from the point of view of management because compensated hemolysis may become “decompensated,” i.e., anemia may suddenly appear in certain circumstances, for instance in pregnancy, folate deficiency, or renal failure interfering with adequate EPO production. Another general feature of chronic HAs is seen when any intercurrent condition, such as an acute infection, depresses erythropoiesis. When this happens, in view of the increased rate of red cell turnover, the effect will be predictably much more marked than in a person who does not have hemolysis. The most dramatic example is infection by parvovirus B19, which may cause a rather precipitous fall in hemoglobin—an occurrence sometimes referred to as *aplastic crisis*.

■ INHERITED HEMOLYTIC ANEMIAS

The red cell has three essential components: (1) hemoglobin, (2) the membrane-cytoskeleton complex, and (3) the metabolic machinery necessary to keep hemoglobin and the membrane-cytoskeleton complex in working order. Diseases caused by inherited abnormalities of hemoglobin, or hemoglobinopathies, are covered in [Chap. 98](#). Here we will deal with diseases of the other two components.

Hemolytic Anemias due to Abnormalities of the Membrane-Cytoskeleton Complex The detailed architecture of the red cell membrane is complex, but its basic design is relatively simple ([Fig. 100-2](#)). The lipid bilayer incorporates phospholipids and cholesterol, and it is spanned by a number of proteins that have their hydrophobic transmembrane domain(s) embedded in the membrane; most of these

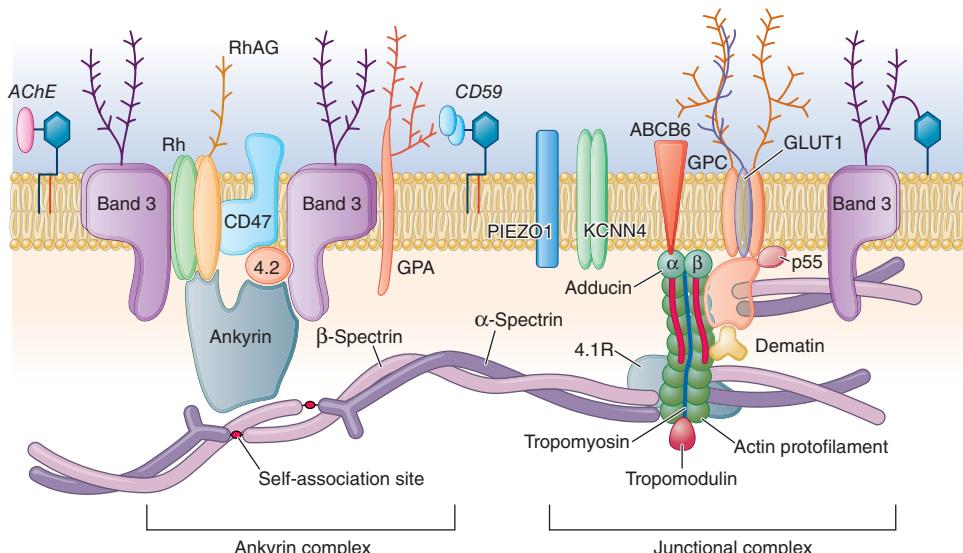


FIGURE 100-2 The red cell membrane and cytoskeleton. Within the membrane lipid bilayer several integral membrane proteins are shown: band 3 (anion exchanger 1 [AE1]) is the most abundant. *PIEZ01* is a mechanoreceptor, *KCNN4*, a Ca^{2+} activated K^+ channel, and *ABCB6* is an ion channel: they are important in the regulation of the red cell volume. Other proteins, e.g., acetylcholinesterase (AChE) and the two complement-regulatory proteins CD59 and CD55, are tethered to the membrane through a glycosylphosphatidylinositol (GPI) anchor: in these cases the entire polypeptide chain is extracellular. Many of the membrane proteins bear polypeptide and/or carbohydrate red cell antigens. Underneath the membrane, the α - β spectrin dimers, that associate head-to-head into tetramers, together with actin and other proteins, form most of the cytoskeleton. The *ankyrin complex*, that also involves the band 4.2 protein, and the *junctional complex*, that involves the band 4.1 protein and dematin, connect the membrane to the cytoskeleton. The ankyrin complex provides mainly radial (also called vertical) connections; the junctional complex provides mainly tangential (also called horizontal) connections: pathogenic changes in the former can cause spherocytosis, whereas pathogenic changes in the latter can cause elliptocytosis; pathogenic changes in spectrin can cause either. Branched lines symbolize carbohydrate moiety of proteins. The various molecules are obviously not drawn to the same scale. Additional explanations are found in the text. (Reproduced with permission from N Young et al: *Clinical Hematology*. Philadelphia, Elsevier, 2006.)

proteins also extend to both the outside (extracellular domains) and the inside of the cell (cytoplasmic domains). Other proteins are tethered to the membrane through a glycosylphosphatidylinositol (GPI) anchor; these have only an extracellular domain. Membrane proteins include energy-dependent ion transporters, ion channels, receptors for complement components, and receptors for other ligands. The most abundant red cell membrane proteins are glycophorins and the so-called band 3, an anion transporter that is an integral membrane protein. The extracellular domains of many of these proteins are heavily glycosylated, and they carry antigenic determinants that correspond to blood groups. Underneath the membrane, and tangential to it, is a network of other proteins that make up the cytoskeleton. The main cytoskeletal protein is the spectrin tetramer, consisting of a head-to-head association of two α -spectrin- β -spectrin heterodimers. The cytoskeleton is linked to the membrane through the *ankyrin complex* (that includes also band 4.2) and the *junctional complex* (that includes adducin and band 4.1) ([Fig. 100-2](#)). These multiprotein complexes make membrane and cytoskeleton intimately connected to each other, thus supporting membrane stability and at the same time providing the erythrocyte with the important property of deformability.

The membrane-cytoskeleton complex has essentially three functions: It is an envelope for the red cell cytoplasm; it maintains the normal red cell shape; it provides cross-membrane transport of electrolytes and of metabolites such as glucose and amino acids. In the membrane-cytoskeleton complex, the individual components are so intimately associated with each other that an abnormality of almost any of them will be disturbing or disruptive, causing mechanical instability of the membrane and/or reduced red cell deformability, ultimately causing hemolysis. These abnormalities are almost invariably inherited mutations; thus diseases of the membrane-cytoskeleton complex belong to the category of inherited HAs. Before the red cells lyse, they often exhibit more or less specific changes that alter the normal biconcave disk shape. Thus, the majority of the diseases in this group have been known for over a century as hereditary spherocytosis (HS) and hereditary elliptocytosis (HE).

More recently a third morphologic entity, whereby on a blood smear the round-shaped central pallor of a red cell is replaced by a linear-shaped central pale area, has earned the name *stomatocytosis*: because this abnormal shape is related to abnormalities of channel molecules, the underlying disorders are also referred to as channelopathies. From an understanding of the molecular basis of these disorders, it has emerged ([Table 100-3](#)) that, although these disorders are predominantly monogenic, no one-to-one correlation exists between a certain gene and a certain disorder. Rather, what has been regarded as a single disorder (e.g., HS) can arise through mutation of one of several genes; conversely, what have been regarded as different disorders can arise through different mutations of the very same gene ([Fig. 100-3](#)).

HEREDITARY SPHEROCYTOSIS This is most common among this group of HAs, with an estimated prevalence of 1:2000–1:5000 in populations of European ancestry. Its identification is credited to Minkowsky and Chauffard, who, at the end of the nineteenth century, reported families who had spherocytes in their peripheral blood ([Fig. 100-4A](#)). In vitro studies revealed that the red cells were abnormally susceptible to lysis in hypotonic media; indeed, the presence of

TABLE 100-3 Inherited Diseases of the Red Cell Membrane-Cytoskeleton Complex

GENE	CHROMOSOMAL LOCATION	PROTEIN PRODUCED	DISEASE(S) WITH CERTAIN MUTATIONS (INHERITANCE)	COMMENTS
<i>SPTA1</i>	1q22-q23	α -Spectrin	HS (recessive) HE (dominant)	Rare Mutations of this gene account for about 65% of HE. More severe forms may be due to coexistence of an otherwise silent mutant allele.
<i>SPTB</i>	14q23-q24.1	β -Spectrin	HS (dominant) HE (dominant)	Rare Mutations of this gene account for about 30% of HE, including some severe forms.
<i>ANK1</i>	8p11.2	Ankyrin	HS (dominant)	May account for majority of HS.
<i>SLC4A1</i>	17q21	Band 3; also known as AE (anion exchanger) or AE1	HS (dominant) Southeast Asia ovalocytosis (dominant) Stomatocytosis (cryohydrocytosis)	Mutations of this gene may account for about 25% of HS. Polymorphic mutation (deletion of nine amino acids); in heterozygotes clinically asymptomatic and protective against <i>Plasmodium falciparum</i> . Certain specific missense mutations shift protein function from anion exchanger to cation conductance.
<i>EPB41</i>	1p33-p34.2	Band 4.1	HE (dominant)	Mutations of this gene account for about 5% of HE, mostly with prominent morphology but little/no hemolysis in heterozygotes; severe hemolysis in homozygotes.
<i>EPB42</i>	15q15-q21	Band 4.2	HS (recessive)	Mutations of this gene account for about 3% of HS.
<i>RHAG</i>	6p21.1-p11	Rhesus-associated glycoprotein	Chronic nonspherocytic hemolytic anemia (recessive)	Very rare; associated with total loss of all Rh antigens. One specific mutation in this gene entails loss of stomatin from the cell membrane, causing overhydrated stomatocytosis.
<i>PIEZ01</i>	16q23-q24	PIEZ01 (mechanosensitive ion channel component 1 channel)	Dehydrated hereditary stomatocytosis (dominant)	Also known as xerocytosis with pseudohyperkalemia. Patients may present with perinatal edema.
<i>KCNN4</i>	19q13.31	KCNN4 Intermediate conductance calcium-activated potassium channel protein 4 (Gardos channel)	Dehydrated hereditary stomatocytosis (dominant)	Clinical presentation similar to that of <i>PIEZ01</i> mutants.
<i>ABCB6</i>	2q35-q36	ATP-binding cassette subfamily B member 6	Familial pseudohyperkalemia (dominant)	Increased potassium leakage upon storage in blood bank condition: this can cause hyperkalemia in the recipient. <i>ABCB6</i> mutation is present in 0.3% of blood donors.
<i>SLC2A1</i>	1p34.2	GLUT1 glucose transporter	Overhydrated hereditary stomatocytosis	Associated with serious neurological manifestations.

Note: *PIEZ01*, *KCNN4*, *ABCB6*, and *GLUT1* are channel molecules; conditions associated with mutations in the respective genes are appropriately named channelopathies.

Abbreviations: HE, hereditary elliptocytosis; HS, hereditary spherocytosis.

osmotic fragility became the main diagnostic test for HS. Today we know that HS, thus defined, is genetically heterogeneous; i.e., it can arise from a variety of mutations in one of several genes (Table 100-3). It has been also recognized that the inheritance of HS is not always autosomal dominant (with the patient being heterozygous); indeed, some of the most severe forms are instead autosomal recessive (with the patient being homozygous).

Clinical Presentation and Diagnosis The spectrum of clinical severity of HS is broad. Severe cases may present in infancy with severe anemia, whereas mild cases may present in young adults or even later in life. The main clinical findings are jaundice, an enlarged spleen, and often gallstones; indeed, it may be the finding of gallstones in a young person that triggers diagnostic investigations.

The variability in clinical manifestations that is observed among patients with HS is largely due to the different underlying molecular lesions (Table 100-3). Not only are mutations of several genes involved, even different mutations of the same gene can give very different clinical manifestations. In milder cases, hemolysis is often compensated (see above), but changes in clinical expression may be seen even in the same patient because intercurrent conditions (e.g., pregnancy, infection) may cause decompensation. The anemia is usually normocytic with the characteristic morphology that gives the disease its name. An increased mean corpuscular hemoglobin concentration (MCHC >34) and increased red cell distribution width (RDW >14%) associated with normal or slightly decreased MCV on an ordinary blood count report should raise the suspicion of HS. The spleen plays a key role in HS

through a dual mechanism. On one hand, like in many other HAs, the spleen itself is a major site of destruction; on the other hand, because the red cells in HS are less deformable, transit through the splenic circulation makes them more prone to vesiculate, thus accelerating their demise.

When there is a family history, it is usually easy to make a diagnosis based on features of HA and typical red cell morphology. However, family history may be negative for at least two reasons. First, the patient may have a *de novo* mutation, i.e., a mutation that has taken place in a germ cell of one of the patient's parents or early after zygote formation. Second, the patient may have a recessive form of HS (Table 100-3). In such cases, more extensive laboratory investigations are required, including osmotic fragility, the acid glycerol lysis test, the eosin-5'-maleimide (EMA)-binding test, and SDS-gel electrophoresis of membrane proteins; these tests are usually carried out in laboratories with special expertise in this area. Sometimes a definitive diagnosis can be obtained only by molecular studies demonstrating a mutation in one of the genes underlying HS (Table 100-3).

TREATMENT

Hereditary Spherocytosis

We do not have a causal treatment for HS; i.e., no way has yet been found to correct the basic defect in the membrane-cytoskeleton structure. Given the special role of the spleen in HS (see above),

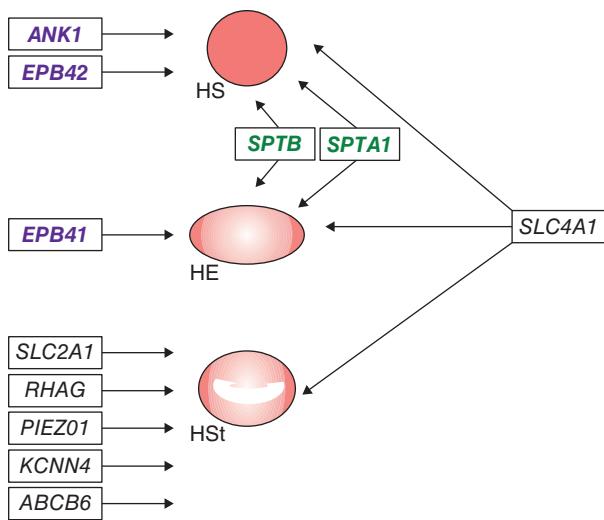


FIGURE 100-3 Hereditary spherocytosis (HS), hereditary elliptocytosis (HE), and hereditary stomatocytosis (HSt) are three morphologically distinct forms of congenital hemolytic anemia. It has emerged that each one can arise from mutation of one of several genes and that different mutations of the same gene can give one or another form. (See also Table 100-3.) Genes encoding membrane proteins are in black; genes encoding cytoskeleton proteins are in green; genes encoding proteins in the junctional and ankyrin complexes are in purple.

splenectomy is often beneficial. Current recommendations are to proceed with splenectomy at the age of 4–6 years in severe cases, to delay splenectomy until puberty in moderate cases, and to avoid splenectomy in mild cases. Partial splenectomy can be considered in certain cases; and it is helpful to know about the outcome of splenectomy in the patient's affected relatives. Before splenectomy, vaccination against encapsulated bacteria (*Neisseria meningitidis* and *Streptococcus pneumoniae*) is imperative; penicillin prophylaxis after splenectomy is controversial. Along with splenectomy, cholecystectomy should not be carried out automatically; but it should be carried out, usually by the laparoscopic approach, whenever it is clinically indicated.

HEREDITARY ELLIPTOCYTOSIS HE is at least as heterogeneous as HS, both from the genetic point of view (Table 100-3, Fig. 100-3) and from the clinical point of view. The global incidence of HE is 1:2000–4000 subjects. Again, it is the shape of the red cells (Fig. 100-4B) that gives the name to the condition, but there is no direct correlation between the elliptocytic morphology and clinical severity. In fact, some mild or even asymptomatic cases may have nearly 100% elliptocytes (or ovalocytes). Indeed, the diagnosis of HE is generally incidental, because hemolysis may be compensated and there may be no anemia, although this may become evident in the course of infection. One particular in-frame deletion of nine amino acids in the *SLC4A1* gene encoding band 3 underlies the so-called Southeast Asia ovalocytosis (SAO): it is not a disease, but rather a polymorphism with a frequency of up to 5–7% in certain populations (e.g., Papua New Guinea, Indonesia, Malaysia, Philippines), presumably as a result of malaria selection; it is asymptomatic in heterozygotes and probably lethal in homozygotes. The cases of HE with the most severe HA are those with biallelic mutations of one of the genes involved (see Fig. 100-3), and these are said to have pyropoikilocytosis (HPP): here the instability of the cytoskeleton protein network may result from decreased tetramerization of spectrin dimers. The red cell volume is decreased (MCV: 50–60 fL), and all kinds of bizarre poikilocytes are seen on the blood smear (Fig. 100-4C). HPP patients have splenomegaly and often benefit from splenectomy.

Channelopathies These rare conditions (see Fig. 100-3) are characterized by abnormalities in red cell ion content and alteration of erythrocyte volume. Cation leak can cause hyperkalemia; in some cases, this leak is accelerated in the cold (the resulting spuriously high serum K^+ is then referred to as pseudo-hyperkalemia). The less rare form,

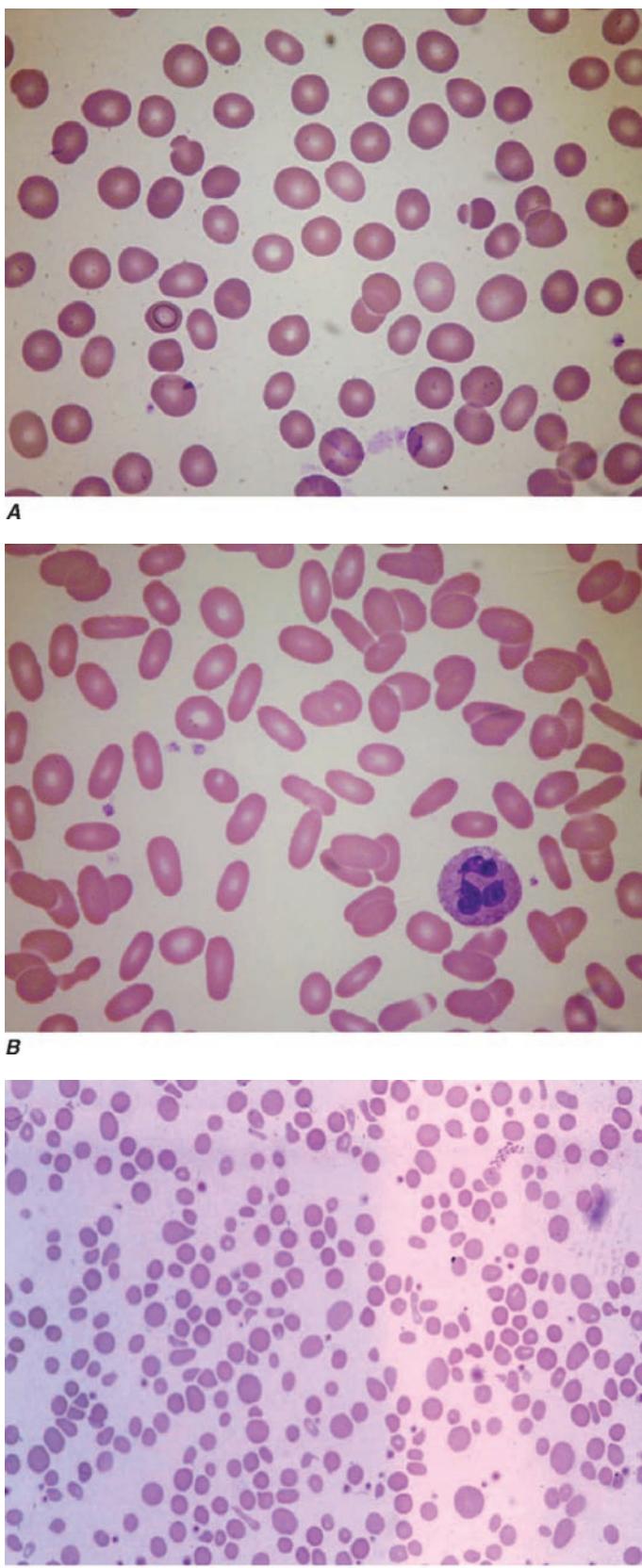


FIGURE 100-4 Peripheral blood smear from patients with membrane-cytoskeleton abnormalities. **A.** Hereditary spherocytosis. **B.** Hereditary elliptocytosis, heterozygote. **C.** Pyropoikilocytosis, with both alleles of the α -spectrin gene mutated.

dehydrated stomatocytosis (DHS; also referred to as xerocytosis) is a (usually compensated) macrocytic hemolytic disorder, with increased MCHC (generally higher than 36 g/dL) associated with mild jaundice. Mutations in either *PIEZ01*, encoding an ion channel activated by

pressure (mechanoreceptor), or in *KCCN4*, encoding the Ca^{2+} activated K^+ channel (Gardos channel) have been recognized to cause DHS (see Table 100-3).

Another form is overhydrated stomatocytosis (OHS): this too is macrocytic ($\text{MCV} > 110 \text{ fL}$), but the MCHC is low ($< 30 \text{ g/dL}$). The underlying mutation is in the Rhesus gene *RHAG*, which encodes an ammonia channel. Yet other patients with stomatocytosis (Table 100-3) have mutations in *SLC4A1* (encoding band 3) and *SLC2A1* (encoding the glucose transporter GLUT1). Mutations of the latter are responsible for *cryohydrocytosis*, a channelopathy in which the red cells swell and burst when they are cooled. In vivo hemolysis can vary from relatively mild to quite severe. Familial hyperkalemia has been recently linked to mutations in *ABCB6*, resulting in abnormal cation leak with extracellular release of a large amount of K^+ (hyperkalemia). Mutations in *ABCB6* have been identified in almost 0.3% of blood donors. However, splenectomy is contraindicated in stomatocytosis due to the significant proportion of severe thromboembolic complications observed in splenectomized DHS patients.

A specialized technique to measure erythrocyte deformability through laser diffraction analysis is *ektacytometry*: this has been used extensively in order to investigate membrane-cytoskeleton abnormalities. For diagnostic purposes, systematic sequencing of a panel of genes in patients' DNA is a powerful approach already in use and destined to be used increasingly.

Enzyme Abnormalities When an important defect in a component of the membrane-cytoskeleton complex is present, hemolysis is a direct consequence of the fact that the very structure of the red cell is compromised. Instead, when one of the enzymes is defective, the consequences will depend on the precise role of that enzyme in the metabolic machinery of the red cell. This machinery has two main

functions: (1) to provide energy in the form of ATP, and (2) to prevent oxidative damage to hemoglobin and to other proteins by providing sufficient reductive potential; the key molecule for this is NADPH.

ABNORMALITIES OF THE GLYCOLYTIC PATHWAY Because red cells, in the course of their differentiation, have sacrificed not only their nucleus and their ribosomes but also their mitochondria, they rely exclusively on the anaerobic portion of the glycolytic pathway for producing ATP, most of which is required by the red cell for cation transport against a concentration gradient across the membrane. If this fails due to a defect of any of the enzymes of the glycolytic pathway (Table 100-4), the result will be hemolytic disease.

Pyruvate Kinase Deficiency Abnormalities of the glycolytic pathway are all inherited and all rare. Among them, deficiency of pyruvate kinase (PK) is the least rare, with an estimated prevalence in most populations of 1:10,000. However, recently, a polymorphic PK mutation (E277K) was found in some African populations with heterozygote frequencies of 1–7%, suggesting that this may be another malaria-related polymorphism. HA secondary to PK deficiency is an autosomal recessive disease (Fig. 100-5).

The clinical picture of homozygous (or biallelic) PK deficiency is that of an HA that often presents in the newborn with neonatal jaundice, requiring nearly always phototherapy and frequently exchange transfusion; the jaundice persists, and it is often associated with reticulocytosis. The anemia is of variable severity; sometimes it is so severe as to require regular blood transfusion treatment, whereas sometimes it is mild, bordering on a nearly compensated hemolytic disorder. As a result, the diagnosis may be delayed: in some cases, it is made, for instance, in a young woman during her first pregnancy, when the anemia may get worse. The delay in diagnosis may be caused in part by the fact that the anemia is often remarkably well tolerated because

TABLE 100-4 Red Cell Enzyme Abnormalities Causing Hemolysis

	ENZYME (ACRONYM)	GENE SYMBOL; CHROMOSOMAL LOCATION	PREVALENCE OF ENZYME DEFICIENCY (RANK)	CLINICAL MANIFESTATIONS EXTRA-RED CELL	COMMENTS
Glycolytic Pathway					
	Hexokinase (HK)	<i>HK1</i> ; 10q22	Very rare		May benefit from splenectomy; BMT ^c
	Glucose 6-phosphate isomerase (G6PI)	<i>GPI</i> ; 19q31.1	Rare (4); at least 60 cases reported ^a	NM, CNS	May benefit from splenectomy
	Phosphofructokinase (PFK) ^b	<i>PFKM</i> ; 12q13	Very rare	Myopathy; myoglobinuria	
	Aldolase	<i>ALDOA</i> ; 16q22-24	Very rare	Myopathy	
	Triose phosphate isomerase (TPI)	<i>TPI1</i> ; 12p13.31	Very rare	CNS (severe), NM	
	Glyceraldehyde 3-phosphate dehydrogenase (GAPD)	<i>GAPDH</i> ; 12p13.31	Very rare	Myopathy	
	Bisphosphoglycerate mutase (BPGM)	<i>BPGM</i> ; 7q33	Very rare		Erythrocytosis rather than hemolysis; some of the rare mutations are in the enzyme active site
	Phosphoglycerate kinase (PGK)	<i>PGK1</i> ; Xq21.1	Very rare	CNS, NM	May benefit from splenectomy; BMT ^c
	Pyruvate kinase (PK)	<i>PKLR</i> ; 1q22	Rare (2) ^a		May benefit from splenectomy; BMT ^c
Redox					
	Glucose 6-phosphate dehydrogenase (G6PD)	<i>G6PD</i> ; Xq28	Common (1) ^a	Very rarely granulocytes	In almost all cases, only AHA from exogenous trigger
	Glutathione synthase	<i>GSS</i> ; 20q11.22	Very rare	CNS	
	Glutathione reductase	<i>GSR</i> ; 8p12	Very rare	Cataracts	AHA from exogenous trigger (favism)
	γ -Glutamylcysteine synthase	<i>GCLC</i> ; 6p12.1	Very rare	CNS	Mutations affect catalytic subunit
	Cytochrome b5 reductase	<i>CYB5R3</i> ; 22q13.2	Rare	CNS	Methemoglobinemia rather than hemolysis
Nucleotide Metabolism					
	Adenylate kinase (AK)	<i>AK1</i> ; 9q34.11	Very rare	CNS	May benefit from splenectomy
	Pyrimidine 5' nucleotidase (P5N)	<i>NTSC3A</i> ; 7p14.3	Rare (3) ^a		May benefit from splenectomy

^aThe numbers from (1) to (4) indicate the ranking order of these enzymopathies in terms of frequency. ^bPFK deficiency is associated with increased glycogen in muscle, and it is also known as glycogen storage disease type VII or Tarui's disease. ^cOccasional report of successful treatment of the hematologic manifestations by BMT.

Abbreviations: AHA, acquired hemolytic anemia; BMT, bone marrow transplantation; CNS, central nervous system; NM, neuromuscular.

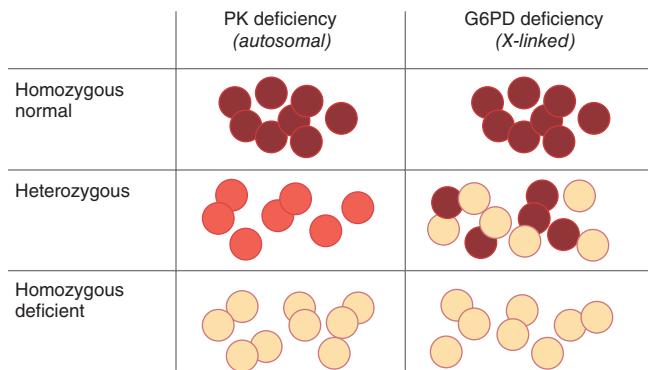


FIGURE 100-5 Different phenotypes of heterozygotes for red cell enzymopathies.

In a heterozygote for deficiency of PK, encoded by an autosomal gene (see Table 100-4), the level of enzyme is about one-half of normal in all red cells. Because this level of enzyme is sufficient, there are no clinical consequences, i.e., PK deficiency is recessive. In a heterozygote for deficiency of G6PD, encoded by an X-linked gene, the situation is quite different: X-chromosome inactivation generates red cell mosaicism, whereby some red cells are entirely normal and others are G6PD deficient. Therefore, G6PD deficiency is expressed in heterozygotes: it is not recessive.

the metabolic block at the last step in glycolysis causes an increase in 2,3-bisphosphoglycerate (or DPG; Fig. 100-1), a major effector of the hemoglobin-oxygen dissociation curve; thus the oxygen delivery to the tissues is enhanced, a remarkable compensatory feat.

TREATMENT

Pyruvate Kinase Deficiency

The management of PK deficiency is mainly supportive. In view of the marked increase in red cell turnover, oral folic acid supplements should be given constantly. Blood transfusion should be used as necessary, and iron chelation may be required even in some patients who, though not receiving blood transfusion, may be developing iron overload (see “General Pathophysiology” above). About one-half of patients sooner or later undergo splenectomy, which usually provides a modest but significant increase in hemoglobin (paradoxically, often reticulocytes also increase considerably). Cholecystectomy may also be required. Some patients with severe disease have received bone marrow transplantation (BMT) from an HLA-identical PK-normal sibling. Prenatal diagnosis has been carried out in a mother who had already had an affected child. A clinical trial of a small molecule that is a specific PK ligand and may increase the stability and/or catalytic efficiency of mutant PK is currently ongoing. Rescue of inherited PK deficiency through lentiviral-mediated human PK gene transfer has been successful in mice. An oral small molecule allosteric activator of PK called *mitapivat* raised hemoglobin levels in about half of PK deficient patients in a small phase 2 study.

Other Glycolytic Enzyme Abnormalities All of these defects are rare to very rare (Table 100-4), and most of them cause HA with varying degrees of severity. It is not unusual for the presentation to be in the guise of severe neonatal jaundice, which may require exchange transfusion; if the anemia is less severe, it may present later in life, or it may even remain asymptomatic and be detected incidentally when a blood count is done for unrelated reasons. The spleen is often enlarged. When other systemic manifestations occur, they can involve the central nervous system (sometimes entailing severe mental retardation, particularly in the case of triose phosphate isomerase deficiency), the neuromuscular system, or both (see Table 100-4). This is not altogether surprising if we consider that these are housekeeping genes, i.e., expressed in all tissues. The diagnosis of HA is usually not difficult, thanks to the triad of normo-macrocytic anemia, reticulocytosis, and hyperbilirubinemia. Enzymopathies should be considered in the differential diagnosis of any chronic Coombs-negative HA. Unlike with membrane disorders, in most

cases of glycolytic enzymopathies, morphologic abnormalities are conspicuous by their absence. A definitive diagnosis can be made only by demonstrating the deficiency of an individual enzyme by quantitative assays; these are carried out in only a few specialized laboratories. If a particular molecular abnormality is already known in the family, then one could test directly for that defect at the DNA level, thus bypassing the need for enzyme assays. Of course the time may be getting nearer when a patient will present with her or his exome already sequenced, and we will need to concentrate on which genes to look up within the file. The principles for the management of these conditions are similar as for PK deficiency. In isolated cases of glycolytic enzyme abnormalities, BMT has been carried out successfully, although unfortunately nonhematologic manifestations, if any, are not reversed.

ABNORMALITIES OF REDOX METABOLISM • Glucose-6-phosphate Dehydrogenase (G6PD) Deficiency G6PD is a housekeeping enzyme critical in the redox metabolism of all aerobic cells (Fig. 100-1). In red cells, its role is even more critical because it is the only source of NADPH, which directly and via glutathione (GSH) defends these cells against oxidative stress (Fig. 100-6). G6PD deficiency-related HA is a prime example of an HA due to interaction between an intracorporeal cause and an extracorporeal cause: indeed, in the vast majority of cases hemolysis is triggered by an exogenous agent. Although the G6PD activity is decreased in most tissues of G6PD-deficient subjects, in other cells the decrease is much less pronounced than in red cells, and it does not seem to impact on clinical expression.

GENETIC CONSIDERATIONS

The G6PD gene is X-linked, and this has important implications. First, because males have only one G6PD gene (i.e., they are hemizygous for this gene), they must be either normal or G6PD deficient. By contrast, females, who have two G6PD genes, can be either normal or deficient (homozygous) or intermediate (heterozygous). Second, as a result of the phenomenon of X chromosome inactivation, heterozygous females are genetic mosaics (see Fig. 100-5), with a highly variable ratio of G6PD-normal to G6PD-deficient cells and an equally variable degree of clinical expression; some heterozygotes can be just as affected as hemizygous males. The enzymatically active form of G6PD is either a dimer or a tetramer of a single protein subunit of 514 amino acids. G6PD-deficient subjects have been found invariably to have mutations in the coding region of the G6PD gene. Almost all of the 230 different mutations known are single missense point mutations, entailing single amino acid replacements in the G6PD protein. In most cases, these mutations cause G6PD deficiency by decreasing the *in vivo* stability of the protein; thus the physiologic decrease in G6PD activity that takes place with red cell aging is greatly accelerated. In some cases, an amino acid replacement can also affect the catalytic function of the enzyme.

Among these mutations, those underlying chronic nonspherocytic hemolytic anemia (CNSHA; see below) are a discrete subset. This much more severe clinical phenotype can be ascribed in some cases to adverse qualitative changes (for instance, a decreased affinity for the substrate glucose-6-phosphate) or simply to the fact that the enzyme deficit is more extreme because of a more severe instability of the enzyme. For instance, a cluster of mutations map at or near the dimer interface, and clearly they compromise severely the formation of the dimer.

Epidemiology G6PD deficiency is widely distributed in tropical and subtropical parts of the world (Africa, Southern Europe, the Middle East, Southeast Asia, and Oceania) (Fig. 100-7) and wherever people from those areas have migrated. A conservative estimate is that at least 500 million people have a G6PD deficiency gene. In several of these areas, the frequency of a G6PD deficiency gene may be as high as 20% or more. It would be quite extraordinary for a trait that causes significant pathology to spread widely and reach high frequencies in many populations without conferring some biologic advantage. Indeed, G6PD is one of the best-characterized examples of genetic polymorphisms in the human species. Clinical field studies and in

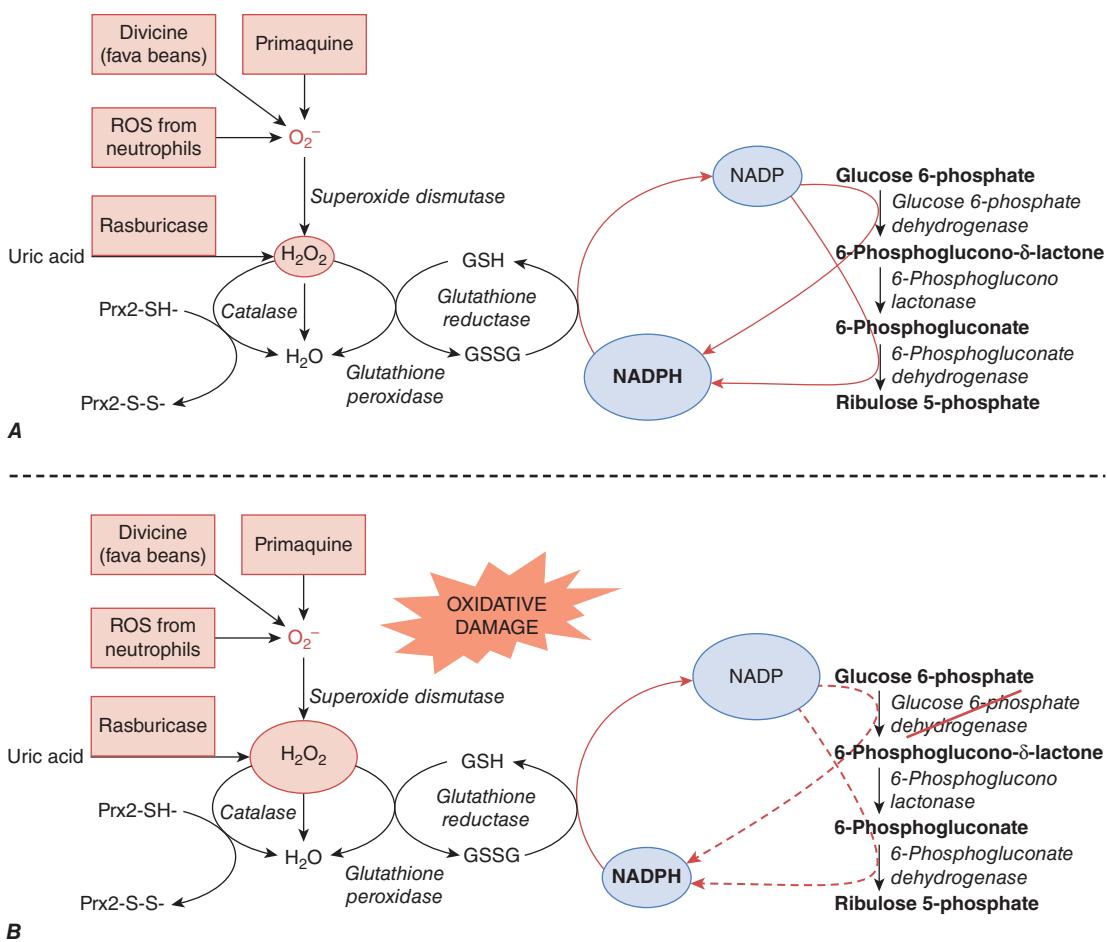


FIGURE 100-6 The role of G6PD in protecting red cells from oxidative damage. **A.** In G6PD-normal red cells, G6PD and 6-phosphogluconate dehydrogenase—two of the enzymes of the pentose phosphate pathway—provide ample supply of NADPH, which in turn regenerates GSH when this is oxidized by reactive oxygen species (e.g., O_2^- and H_2O_2). Thus when O_2^- (meant here to represent itself and other reactive oxygen species or ROS) is produced by pro-oxidant compounds such as primaquine, or the glucosides in fava beans (divicine), or the oxidative burst of neutrophils, these ROS are rapidly neutralized; similarly, when rasburicase administered to degrade uric acid produces an equimolar amount of hydrogen peroxide, this is rapidly degraded by the combined action of glutathione peroxidase, catalase, and Prx2 (peroxiredoxin-2: all three mechanisms are NADPH dependent). **B.** In G6PD-deficient red cells, where the enzyme activity is reduced, NADPH production is limited, and it may not be sufficient to cope with the excess ROS generated by pro-oxidant compounds, and the consequent excess hydrogen peroxide. This diagram also explains why a defect in glutathione reductase has very similar consequences to G6PD deficiency.

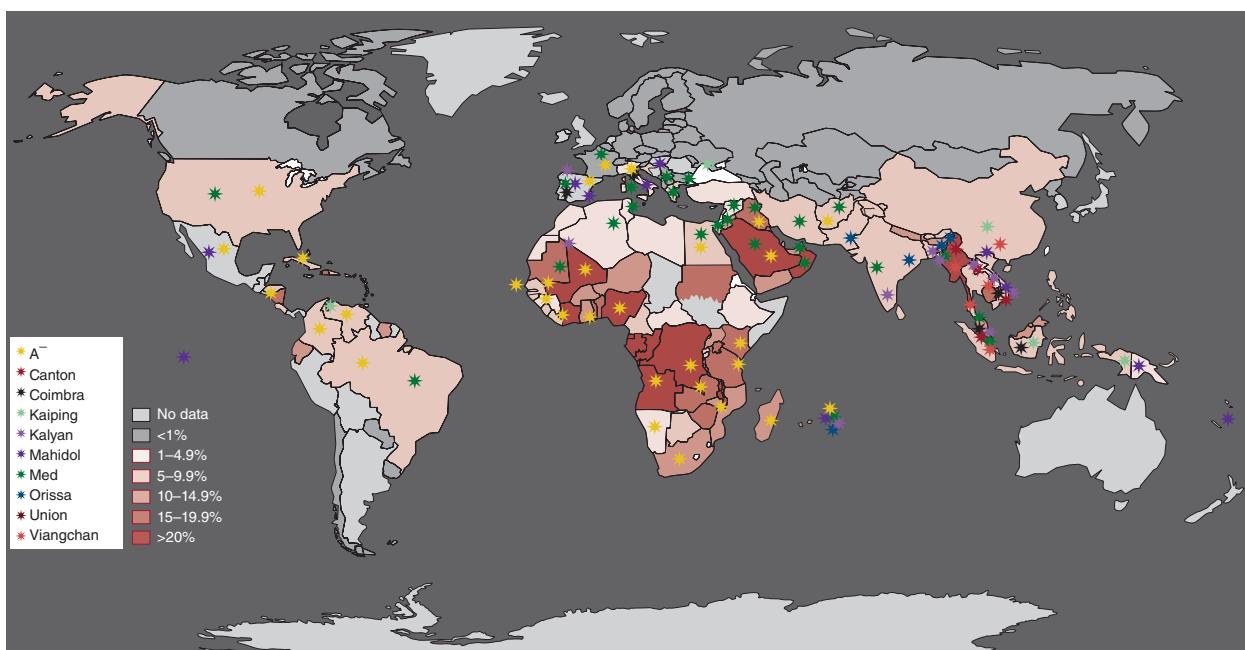


FIGURE 100-7 Epidemiology of G6PD deficiency throughout the world. Each country on the map is shaded in a color based on the best estimate of the mean frequency of G6PD deficiency allele(s) in that country (this is the same as the frequency of G6PD deficient males). The small panel on the left gives the key to color shadings corresponding to each country. The larger panel gives a color-coded list of ten common G6PD variants associated with G6PD deficiency: asterisk-shaped symbols in the corresponding colors are shown in the countries where these variants have been observed (for graphic reasons symbols could not be inserted in all countries). (Republished with permission of American Society of Hematology, from *Glucose-6-phosphate dehydrogenase deficiency*, L Luzzatto et al. 136:1225, 2020 permission conveyed through Copyright Clearance Center, Inc.)

vitro experiments strongly support the view that G6PD deficiency has been selected by *Plasmodium falciparum* malaria because it confers a relative resistance against this highly lethal infection. As in other cases of balanced polymorphism, it is heterozygotes, therefore females, who are protected. Different G6PD variants underlie G6PD deficiency in different parts of the world. Examples of widespread variants are G6PD Mediterranean on the shores of that sea, in the Middle East, and elsewhere; G6PD A- in Africa, in the Middle East, and in Southern Europe; G6PD Orissa in India; G6PD Viangchan and G6PD Mahidol in Southeast Asia; G6PD Kaiping and G6PD Canton in China; and G6PD Union worldwide. The heterogeneity of polymorphic G6PD variants is proof of their independent origin, further supporting the notion of selection by a common environmental agent, namely malaria, in keeping with the concept of convergent evolution (Fig. 100-7).

Clinical Manifestations The vast majority of people with G6PD deficiency remain clinically asymptomatic throughout their lifetime; however, all of them have an increased risk of developing neonatal jaundice (NNJ) and a risk of developing acute HA (AHA) when challenged by a number of oxidative agents. NNJ related to G6PD deficiency is rarely present at birth; the peak incidence of clinical onset is between day 2 and day 3, and in most cases the anemia is not severe. However, NNJ can be very severe in some G6PD-deficient babies, especially in association with prematurity, infection, and/or environmental factors (such as naphthalene-camphor balls, which may be used in babies' bedding and clothing); and the risk of severe NNJ is also increased by the coexistence of a monoallelic or biallelic mutation in the uridyl transferase gene (*UGT1A1*; the same mutations are associated with Gilbert's syndrome). It is imperative to manage promptly NNJ associated with G6PD deficiency, because it can produce kernicterus and permanent neurologic damage.

AHA can develop as a result of three types of triggers: (1) fava beans, (2) infections, and (3) drugs (Table 100-5). Typically, a hemolytic attack starts with malaise, weakness, and abdominal or lumbar pain. Within a timeframe of several hours to 2–3 days, the patient develops jaundice and often dark urine. The onset can be extremely abrupt, especially with favism in children. The anemia is moderate to extremely severe, usually normocytic and normochromic, and due partly to intravascular hemolysis; hence, it is associated with hemoglobinuria, hemoglobinuria, high LDH, and low or absent plasma haptoglobin. The blood film shows anisocytosis, polychromasia, and spherocytes; in addition, the most typical feature of G6PD deficiency is the presence of bizarre

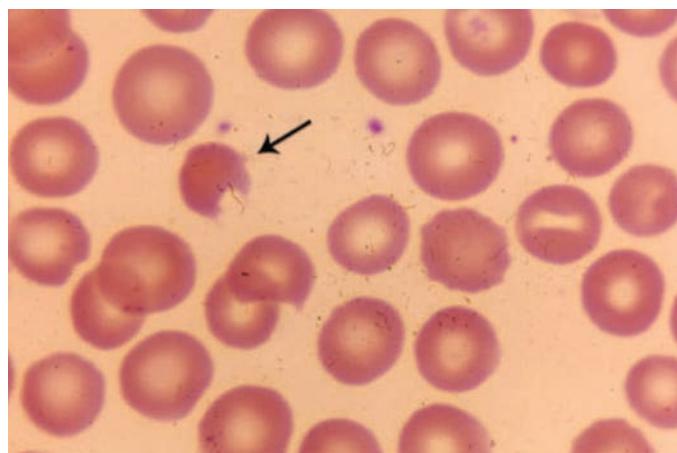


FIGURE 100-8 Peripheral blood smear from a glucose-6-phosphate dehydrogenase (G6PD)-deficient boy experiencing hemolysis. Note the red cells that are misshapen and called "bite" cells. (From MA Lichtman et al: *Lichtman's Atlas of Hematology*: <http://www.accessmedicine.com>. Copyright © The McGraw-Hill Companies, Inc. All rights reserved.)

poikilocytes, with red cells that appear to have unevenly distributed hemoglobin ("hemighosts") and red cells that appear to have had parts of them bitten away ("bite cells" or "blister cells") (Fig. 100-8). A classical test, now rarely carried out, is supravitral staining with methyl violet, which, if done promptly, reveals the presence of Heinz bodies (consisting of precipitates of denatured hemoglobin and hemichromes), which are regarded as a signature of oxidative damage to red cells (they are also seen with unstable hemoglobins). Since there is also a substantial component of extravascular hemolysis, unconjugated bilirubin is high and there is often clinical icterus. The most serious threat from AHA in adults is the development of acute renal failure (this is exceedingly rare in children). Once the threat of acute anemia is over and in the absence of comorbidity, full recovery from AHA associated with G6PD deficiency is the rule.

It was primaquine (PQ)-induced AHA that led to the discovery of G6PD deficiency, but this drug has not been very prominent subsequently because it is not necessary for the treatment of life-threatening *P. falciparum* malaria. Today there is a revival in the use of PQ for two reasons. First, it is the only effective agent for eliminating the gametocytes of *P. falciparum* (thus preventing further transmission): a small single dose (0.25 mg/kg) is required, and it is safe for G6PD-deficient persons. Second, a 14-day course of PQ is needed for eliminating the hypnozoites of *Plasmodium vivax* (thus preventing endogenous relapse). In countries aiming to eliminate malaria, there may be a call for mass administration of PQ; this ought to be associated with G6PD testing. At the other end of the historic spectrum, the latest additions to the list of potentially hemolytic drugs (Table 100-5) are rasburicase and pegloticase; again G6PD testing ought to be made mandatory before giving either of these drugs, because fatal cases have been reported upon using one of these drugs, which generate hydrogen peroxide, in newborns with kidney injury and in adults with tumor lysis syndrome.

Although drug-induced AHA has been prominent in the study of G6PD deficiency, the most common clinical manifestations are in fact NNJ and favism, both of which are of public health importance in many populations. Contrary to beliefs that are still widespread, fava bean pollen inhalation does not cause favism, and other beans are safe.

A very small minority of subjects with G6PD deficiency have CNSHA of variable severity. The patient is nearly always a male, usually with a history of NNJ, who may present with anemia, unexplained jaundice, or gallstones later in life. The spleen may be enlarged. The severity of anemia ranges in different patients from borderline to transfusion dependent. The anemia is usually normo-macrocyclic, with reticulocytosis. Bilirubin and LDH are increased. Although hemolysis is, by definition, chronic in these patients, they are also vulnerable to acute oxidative damage, and therefore the same agents that can cause AHA in people with the ordinary type of G6PD deficiency will cause

TABLE 100-5 Drugs That Carry Risk of Clinical Hemolysis in Persons with Glucose 6-Phosphate Dehydrogenase Deficiency

	DEFINITE RISK	POSSIBLE RISK	DOUBTFUL RISK
Antimalarials	Primaquine	Chloroquine; hydroxychloroquine	Quinine
	Dapsone/ chlorproguanil ^a		
Sulphonamides/ sulphones	Dapsone	Sulfamethoxazole Sulfasalazine Sulfadimidine	Sulfisoxazole Sulfadiazine
Antibacterial/ antibiotics	Cotrimoxazole Nalidixic acid	Ciprofloxacin Norfloxacin	Chloramphenicol <i>p</i> -Aminosalicylic acid
	Nitrofurantoin Niridazole		
Antipyretic/ analgesics	Acetanilide	Acetylsalicylic acid high dose (>3 g/d)	
	Phenazopyridine		Acetaminophen Phenacetin
Other	Rasburicase Naphthalene Methylene blue	Vitamin K analogues Ascorbic acid (>1 g)	Doxorubicin Probenecid

^aMarketed as Lapdap from 2003 to 2008.

severe exacerbations in people with CNSHA associated with G6PD deficiency. In some cases of CNSHA, the deficiency of G6PD is so severe in granulocytes that it limits their capacity to produce an oxidative burst, with consequent increased susceptibility to some bacterial infections.

Laboratory Diagnosis The suspicion of G6PD deficiency can be confirmed by semiquantitative methods often referred to as screening tests, which are suitable for population studies and can correctly classify male subjects, in the steady state, as G6PD normal or G6PD deficient. However, in clinical practice, a diagnostic test is usually needed when the patient has had a hemolytic attack: whereby the oldest, most G6PD-deficient red cells have been selectively destroyed, and young red cells, having higher G6PD activity, are being released into the circulation. Under these conditions, only a quantitative test can give a definitive result. In males, this test will identify normal hemizygotes and G6PD-deficient hemizygotes; among females, some heterozygotes will be missed, but those who are at most risk of hemolysis will be identified. Of course, G6PD deficiency also can be diagnosed by DNA testing. Currently easy-to-use “point of care” tests for G6PD deficiency are becoming available, geared especially to the prospect of mass administration of PQ or of the newly introduced derivative tafenoquine.

TREATMENT

G6PD Deficiency

The AHA of G6PD deficiency is largely preventable by avoiding exposure to triggering factors of previously screened subjects. Of course, the practicability and cost-effectiveness of screening depend on the prevalence of G6PD deficiency in each individual community. Favism is entirely preventable in G6PD-deficient subjects by not eating fava beans. Drug-induced hemolysis can be prevented by testing for G6PD deficiency before prescribing; in many cases one can use alternative drugs. When AHA develops and once its cause is recognized, no specific treatment is needed in most cases. However, if the anemia is severe, it may be a medical emergency, especially in children, requiring immediate action, including blood transfusion. This has been the case with an antimalarial drug combination containing dapsone (called Lapdap, introduced in 2003) that has caused severe acute hemolytic episodes in children with malaria in several African countries; after a few years, the drug was taken off the market. If there is acute renal failure, hemodialysis may be necessary, but if there is no previous kidney disease, recovery is the rule. The management of NNJ associated with G6PD deficiency is no different from that of NNJ due to other causes.

In cases with CNSHA, if the anemia is not severe, regular folic acid supplements and regular hematologic surveillance will suffice. It will be important to avoid exposure to potentially hemolytic drugs, and blood transfusion may be indicated when exacerbations occur, mostly in concomitance with intercurrent infection. In rare patients, regular blood transfusions may be required, in which case appropriate iron chelation should be instituted. Unlike in HS, there is no evidence of selective red cell destruction in the spleen; however, in practice, splenectomy has proven beneficial in severe cases.

Other Abnormalities of the Redox System As mentioned previously, GSH is a key player in the defense against oxidative stress. Inherited defects of GSH metabolism are exceedingly rare, but each one can give rise to chronic HA (Table 100-4). A rare, peculiar, and severe but usually self-limited HA occurring in the first month of life, called *infantile poikilocytosis*, may be associated with deficiency of glutathione peroxidase (GSHPX) due not to an inherited abnormality, but to transient nutritional deficiency of selenium, an element essential for the activity of GSHPX.

PYRIMIDINE 5'-NUCLEOTIDASE (PSN) DEFICIENCY PSN is a key enzyme in the catabolism of nucleotides arising from the degradation of nucleic acids that takes place in the final stages of erythroid cell maturation. How exactly its deficiency causes HA is not well understood,

but a highly distinctive feature of this condition is a morphologic abnormality of the red cells known as *basophilic stippling*. The condition is rare, but it probably ranks third in frequency among red cell enzyme defects (after G6PD deficiency and PK deficiency). The anemia is lifelong, of variable severity, and may benefit from splenectomy.

Familial (Atypical) Hemolytic-Uremic Syndrome (aHUS)

This term is used to designate a group of rare disorders, mostly affecting children, characterized by microangiopathic HA with presence of fragmented erythrocytes in the peripheral blood smear, thrombocytopenia (usually mild), and acute renal failure. (The word *atypical* in this phrase should be consigned to history: it was introduced originally to distinguish this condition from the hemolytic-uremic syndrome [HUS] caused by infection with *Escherichia coli* producing the Shiga toxin, regarded as *typical*.) The genetic basis of atypical HUS (aHUS) has been elucidated. Studies of >100 families have revealed that those family members who developed HUS had mutations in any one of several genes encoding complement regulatory proteins: complement factor H (CFH), CD46 or membrane cofactor protein (MCP), complement factor I (CFI), complement component C3, complement factor B (CFB), thrombomodulin, and others. Thus, whereas all other inherited HAs are due to intrinsic red cell abnormalities, this group is unique in that hemolysis results from an inherited defect external to red cells (Table 100-1). Because the regulation of the complement cascade has considerable redundancy, in the steady state any of the above abnormalities can be tolerated. However, when an intercurrent infection or some other trigger briskly activates complement the deficiency of one of the complement regulators becomes critical. Endothelial cells get damaged, especially in the kidney; at the same time, and partly as a result of this, there will be brisk hemolysis (thus, the more common Shiga toxin-related HUS (Chap. 166) can be regarded as a phenotype of aHUS). aHUS is a severe disease, with up to 15% mortality in the acute phase and up to 50% of cases progressing to end-stage renal disease (ESRD). Not infrequently, aHUS undergoes spontaneous remission. Because it is an inherited abnormality, it is not surprising that, given renewed exposure to a trigger, the syndrome will tend to recur; when it does, the prognosis is always serious. The traditional treatment has been plasma exchange, which will supply the deficient complement regulator. This has changed since the introduction of the anti-C5 complement inhibitor eculizumab (see “Paroxysmal Nocturnal Hemoglobinuria”) was found to greatly ameliorate the microangiopathic picture, with improvement in platelet counts and in renal function, thus abrogating the need for plasma exchange, which is not always effective and not free of complications. Because the basis of aHUS is genetic, and relapses are always possible even after complete remission, there is a rationale for continuing eculizumab indefinitely, especially in order to prevent ESRD. Patients who relapsed after discontinuing eculizumab have responded again. Discontinuation of eculizumab might be reasonable especially in patients heterozygous for a MCP mutation. However, there is no evidence base at the moment for balancing the pros and cons of lifetime eculizumab (a very expensive drug).

ACQUIRED HEMOLYTIC ANEMIA

Mechanical Destruction of Red Cells Although red cells are characterized by the remarkable deformability that enables them to squeeze through capillaries narrower than themselves for thousands of times in their lifetime, there are at least two situations in which they succumb to shear, if not to wear and tear; the result is intravascular hemolysis, resulting in hemoglobinuria (Table 100-6). One situation is acute and self-inflicted, *march hemoglobinuria*. Why sometimes a marathon runner may develop this complication, whereas on another occasion, this does not happen, we do not know (perhaps her or his footwear needs attention). A similar syndrome may develop after prolonged barefoot ritual dancing or intense playing of bongo drums. The other situation is chronic and iatrogenic (it has been called *microangiopathic hemolytic anemia*). It takes place in patients with prosthetic heart valves, especially when paraprosthetic regurgitation is present. If the hemolysis consequent on mechanical trauma to the red cells is mild, and if the supply of iron is adequate, the loss may be largely

TABLE 100-6 Diseases and Clinical Situations in Which Hemolysis Is Largely Intravascular

	ONSET/TIME COURSE	MAIN MECHANISM	APPROPRIATE DIAGNOSTIC PROCEDURE	COMMENTS
Mismatched blood transfusion	Abrupt	Nearly always ABO incompatibility	Repeat cross-match	
Paroxysmal nocturnal hemoglobinuria (PNH)	Chronic with acute exacerbations	Complement (C)-mediated destruction of CD59(-) red cells	Flow cytometry to display a CD59(-) red cell population	Exacerbations due to C activation through any pathway
Paroxysmal cold hemoglobinuria (PCH)	Acute	Immune lysis of normal red cells	Test for Donath-Landsteiner antibody	Often triggered by viral infection
Septicemia	Very acute	Exotoxins produced by <i>Clostridium perfringens</i>	Blood cultures	Other organisms may be responsible
Microangiopathic	Acute or chronic	Red cell fragmentation	Red cell morphology on blood smear	Different causes ranging from endothelial damage to hemangioma to leaky prosthetic heart valve
March hemoglobinuria	Abrupt	Mechanical destruction	Targeted history taking	Has been reported after extreme ritual dancing
Favism	Acute	Destruction of older fraction of G6PD-deficient red cells	G6PD assay	Triggered by ingestion of large dish of fava beans ^a

^aThe trigger of acute hemolytic anemia, often with hemoglobinuria, can be infection or a drug (see Table 100-5) rather than fava beans. Hemoglobinuria may or may not be reported by patient; but it is often macroscopic, i.e., recognizable by simple inspection of urine.

Abbreviation: G6PD, glucose 6-phosphate dehydrogenase.

compensated; if more than mild anemia develops, reintervention to correct regurgitation may be required.

Infection By far the most frequent infectious cause of HA in endemic areas is malaria (**Chap. 224**). In other parts of the world, the most frequent direct cause is probably Shiga toxin-producing *E. coli* O157:H7, now recognized as the main etiologic agent of HUS, which is more common in children than in adults (**Chap. 161**). Life-threatening intravascular hemolysis, due to a toxin with lecithinase activity, occurs with *Clostridium perfringens* sepsis, particularly following open wounds, septic abortion, or as a disastrous accident due to a contaminated blood unit. Rarely, and if at all in children, HA is seen with sepsis or endocarditis from a variety of organisms. In addition, bacterial and viral infections can cause HA by indirect mechanisms (see Table 100-6).

Immune Hemolytic Anemias These can arise through at least two distinct mechanisms. First, when an antibody directed against a certain molecule (e.g., a drug) reacts with that molecule, red cells may get caught in the reaction (the so-called innocent bystander mechanism: see section below on Hemolytic Anemia from Toxic Agents and Drugs), whereby they are damaged or destroyed. Second, and more frequently, a true autoantibody is directed against a red cell antigen, i.e., a molecule present on the surface of red cells. Autoimmune hemolytic anemias have been originally classified into two types, depending on the thermal amplitude of the autoantibodies involved: this classification is valid, because the two types have different pathophysiological and clinical features.

AUTOIMMUNE HEMOLYTIC ANEMIA, WARM TYPE (WAIHA: FOR SIMPLICITY WE WILL USE THE ACRONYM AIHA) This type has an estimated incidence in the United States of about 1–3:100,000 per year, and a prevalence of 17:100,000. AIHA can be serious since even with appropriate management the mortality is of the order of 5–10%.

Clinical Features and Diagnosis The onset is often abrupt and can be dramatic. The hemoglobin level may drop, within days, to as low as 4 g/dL; the massive red cell removal will produce jaundice, and sometimes the spleen is enlarged. When this triad is present, the suspicion of AIHA must be high. The reticulocyte count is typically elevated, except when erythroid precursors are also targeted by the autoantibody attack. LDH may also be elevated. In some cases, AIHA can be associated, on first presentation or subsequently, with autoimmune thrombocytopenia. This double autoimmune condition, referred to as Evans syndrome, may be a manifestation of common variable immune deficiency, and in children it may suggest one of several primary immune deficiency syndromes. Evans syndrome signals high-risk disease. Other predictors of the outcome and of the probability of relapse of AIHA are severe

anemia (Hb <6 g/dL), certain characteristics of the antibody, acute renal failure, and infection.

There are few situations in hematology where one laboratory test is as informative as the direct antiglobulin test developed in 1945 by R. R. A. Coombs and known since then by this name. The currently recommended version of this test uses in the first instance a “broad-spectrum” reagent, i.e., one that will detect not only immunoglobulins (Ig) but also complement (C) components (usually C3 fragments) bound to the surface of the patient’s red cells. If the test is positive (and barring special circumstances such as previous blood transfusion), it is practically diagnostic of AIHA, and one can then determine, by using specific reagents, whether Ig or C or both are implicated. The sensitivity of the Coombs test varies depending on the techniques that are used: in general, the test is positive if there are an average of at least 400 molecules of Ig and/or C on each red cell; but with more advanced techniques involving flow cytometry analysis or enzyme-linked radiolabeled tests allowing the detection of ~30–40 antibody molecules per erythrocyte, the sensitivity can be pushed to as low as 30–40 molecules per red cell. Therefore liaison with a specialized laboratory is desirable; a dual direct antiglobulin test has also been developed. In the past the diagnosis of “Coombs-negative AIHA” was regarded as a last resort, but it is important to know that a patient with this label may have severe AIHA, because if the antibody is powerful (high affinity/avidity), few molecules may be sufficient to opsonize red cells. Based on the Coombs test findings as well as on the thermal characteristics and the antigenic specificities of the autoantibodies (**Table 100-7**), AIHA has been classified into subtypes.

In AIHA the autoantibody reacts best at 37°C and it is usually Rhesus-specific (sometimes specifically anti-e). The main mechanism of hemolysis in AIHA is that the Fc portion of the IgG antibody bound to red cells is recognized by the Fc receptor of macrophages: this will trigger erytrophagocytosis wherever macrophages are abundant, i.e., in the liver, in the bone marrow, but especially in red pulp of the spleen (see **Fig. 100-9**) that, also because of its special anatomy, is often the predominant site of red cell destruction.

AIHA may be seen in isolation (and it is then called *idiopathic*) or as secondary to other disorders such as systemic autoimmune disorders (systemic lupus erythematosus [SLE]: sometimes AIHA may be the first manifestation that leads to a diagnosis of SLE) or lymphoproliferative disorders (Table 100-7). Like all autoimmune diseases, AIHA must arise from a dysregulation of immunity. It is therefore not surprising that it is increasingly being recognized in chronic lymphocytic leukemia (CLL), whether treated or untreated; after BMT; and after solid organ transplantation entailing immunosuppressive treatment. Recently, warm antibody AIHA has also occurred as a side effect of the use of immune checkpoint inhibitors, such as nivolumab, in patients with various types of cancer.

CLINICAL SETTING	TYPE OF ANTIBODY	
	COLD, MOSTLY IgM, OPTIMAL TEMPERATURE 4°C–30°C	WARM, MOSTLY IgG, OPTIMAL TEMPERATURE 37°C; OR MIXED
Primary	CAD	AIHA (idiopathic)
Secondary to viral infection	EBV CMV Other	Parvovirus B19 HIV HCV EBV Viral vaccines
Secondary to other infection	Mycoplasma infection: paroxysmal cold hemoglobinuria	Babesia
Secondary to/ associated with other disease	CAD in: Waldenström's disease Lymphoma	AIHA in: SLE, scleroderma, RA CLL Lymphoproliferative disorders Multiple myeloma Other malignancy Chronic inflammatory disorders (e.g., IBD) Thyroiditis (including Hashimoto) After allogeneic HSCT Common variable immunodeficiency After immune checkpoint modulating drugs
Secondary to drugs: drug-induced immune hemolytic anemia	Small minority (e.g., with lenalidomide)	Majority: currently most common culprit drugs are cefotetan, ceftriaxone, piperacillin, methydopa, fludarabine
		Drug-dependent: antibody destroys red cells only when drug present (e.g., rarely penicillin)
		Drug-independent: antibody can destroy red cells even when drug no longer present (e.g., methydopa)
Associated with	Pregnancy	

Abbreviations: AIHA, autoimmune hemolytic anemia; CAD, cold agglutinin disease; CLL, chronic lymphocytic leukemia; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; HSCT, hematopoietic stem cell transplantation; IBD, inflammatory bowel disease; SLE, systemic lupus erythematosus; RA: rheumatoid arthritis.

TREATMENT

Warm Antibody Autoimmune Hemolytic Anemia

Severe acute AIHA can be a medical emergency. The immediate treatment almost invariably includes transfusion of red cells. This may pose a special problem because many or all of the blood units cross-matched may be incompatible. In these cases, it is often correct, if paradoxical, to transfuse ABO-matched but incompatible blood: the rationale being that the transfused red cells will be destroyed no less—but no more—than the patient's own red cells, and in the meantime the patient stays alive. A situation like this requires close liaison and understanding between the clinical unit treating the patient and the blood transfusion/serology lab. Whenever the anemia is not immediately life-threatening, blood transfusion should be withheld (because compatibility problems may increase with each unit of blood transfused), and medical treatment started immediately with prednisone (1 mg/kg per day), which will produce a remission promptly in at least one-half of patients. Rituximab (anti-CD20), previously regarded as second-line treatment, is increasingly being used at a relatively low dose (100 mg/week × 4), together with prednisone as part of first-line treatment. It is especially encouraging that this approach seems to reduce the rate of relapse, a common occurrence in AIHA.

For patients who do relapse or are refractory to medical treatment, additional therapeutic strategies are now available. Splenectomy does not cure the disease, but it can produce significant benefit by removing a major site of hemolysis, thus improving the anemia and/or reducing the need for other therapies (e.g., the dose of prednisone); of course, splenectomy is not free of risk, as it entails increased risk of sepsis and of thrombosis. The response rate to splenectomy and to rituximab are similar. Since the introduction of rituximab, azathioprine, cyclophosphamide, cyclosporine, mycophenolate and intravenous immunoglobulin have become second- or third-line agents. In very rare severe refractory cases, one may have to consider a high dose of cyclophosphamide (50 mg/kg/d for 4 days) followed by a myelo-stimulating agent to support bone marrow or the anti-CD52 agent, alemtuzumab. When severe anemia is associated with reticulocytopenia, the use of erythropoietin may help to reduce or avoid the requirement for transfusion of red cells.

PAROXYSMAL COLD HEMOGLOBINURIA (PCH) PCH is a rare form of AIHA occurring mostly in children, usually triggered by a viral infection, usually self-limited, and characterized by the so-called Donath-Landsteiner antibody. In vitro, this antibody has unique serologic features; it has usually anti-P specificity and it binds to red cells only at a low temperature (optimally at 4°C), but when the temperature is shifted to 37°C, lysis of red cells takes place in the presence of complement. Consequently, in vivo there is intravascular hemolysis, resulting in hemoglobinuria. Clinically the differential diagnosis must include other causes of hemoglobinuria (Table 100-6), but the presence of the Donath-Landsteiner antibody will prove PCH. Active supportive treatment, including blood transfusion, may be needed to control the anemia; subsequently, recovery is the rule.

COLD AGGLUTININ DISEASE This designation indicates the other main type of AIHA, which has quite different features when compared with wAIHA. First, cold agglutinin disease (CAD) is a chronic and more frequently indolent condition—in contrast to the abrupt onset of warm antibody AIHA. Second, the term *cold* refers to the fact that the autoantibody involved reacts with red cells poorly or not at all at 37°C, whereas it reacts strongly at lower temperatures. As a result, hemolysis is more prominent the more the body is exposed to the cold. Third, the antibody is produced by a clone of autoreactive B lymphocytes. Sometimes the antibody concentration in the serum is high enough to show up as a spike in plasma protein electrophoresis, thus qualifying CAD as an IgM monoclonal gammopathy; however, it differs from Waldenström macroglobulinemia by not having the characteristic MYD88 mutation (see Chap. 111): there is instead, in the B-cell clone of a majority of CAD patients, a somatic mutation in the KMT2D gene, encoding a lysine histone methylase that seems to favor proliferation. The antibody produced by the B-cell clone is IgM; usually it has an anti-I specificity (the I antigen is present on the red cells of almost everybody), and it may have a very high titer (1:100,000 or more has been observed). IgM, when bound to red cells, is a powerful activator of the complement cascade, with ultimate formation of the membrane attack complex (see Fig. 100-9): this will directly cause destruction of red cells (*intravascular hemolysis*: indeed, CAD patients may present with hemoglobinuria). In addition, once complement is activated C3b will bind to red cells that, thus opsonized, will be destroyed by macrophages (*extravascular hemolysis*): unlike in AIHA, there is no predominance of the spleen in this process.

In mild forms of CAD, avoidance of exposure to cold may be all that is needed to enable the patient to have a reasonably comfortable quality of life; but in more severe forms, the management of CAD is not easy. Plasma exchange will remove antibodies and is, therefore, in theory, a rational approach in severe cases. However, the management of CAD has changed significantly with the advent of the anti-CD20 antibody rituximab: up to 60% of patients respond. If remission is followed by relapse, a new course of rituximab may be again effective, and remissions may be more durable with a rituximab-fludarabine combination, in particular in CAD associated with lymphoproliferative disorders. Therefore, even

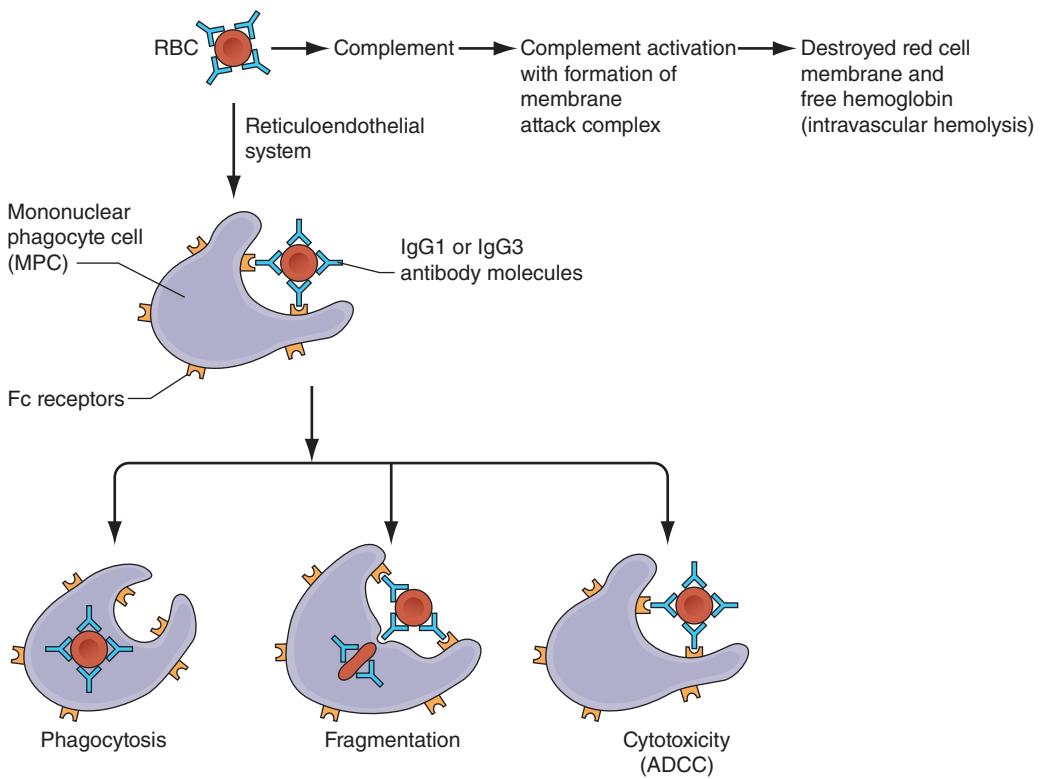


FIGURE 100-9 Mechanism of antibody-mediated immune destruction of red blood cells (RBCs). The three bottom images illustrate three different modalities of extravascular hemolysis. ADCC, antibody-dependent cell-mediated cytotoxicity. (Reproduced with permission from N Young et al: *Clinical Hematology*. Philadelphia, Elsevier, 2006.)

in the absence of a formal trial, rituximab has become de facto first-line treatment: especially since previously used immunosuppressive/cytotoxic agents, although they can reduce the antibody titer, have limited clinical efficacy and, in view of the chronic nature of CAD, their side effects may prove unacceptable. Unlike in AIHA, prednisone and splenectomy are ineffective. In the management of CAD in relapse, there is an emerging role for the B-cell receptor inhibitors venetoclax and ibrutinib, as well as for the proteasome inhibitor bortezomib. A different approach targeting complement inhibitors has been also explored by using eculizumab (anti-C5) or sutimlimab (anti-C1s); a limitation of this approach is that hemolysis will be curbed only for as long as these agents are administered.

In terms of supportive treatment, blood transfusion may be helpful—in spite of the fact that red cells from the donor, being I-positive, will survive no longer than those of the patient: both the blood bag and the patient's extremities must be kept warm during transfusion.

Hemolytic Anemia from Toxic Agents and Drugs A number of chemicals with oxidative potential, whether medicinal or not, can cause hemolysis even in people who are not G6PD deficient (for which, see above). Examples are hyperbaric oxygen (or 100% oxygen), nitrates, chlorates, methylene blue, dapsone, cisplatin, and numerous aromatic (cyclic) compounds. Other chemicals may be hemolytic through nonoxidative, largely unknown mechanisms; examples include arsine, stibine, copper, and lead. The HA caused by lead poisoning is characterized by basophilic stippling; it is in fact a phenocopy of that seen in P5N deficiency (see above), suggesting it is mediated at least in part by lead inhibiting this enzyme.

In these cases, hemolysis appears to be mediated by a direct chemical action on red cells. But drugs can cause hemolysis through at least two other mechanisms. (1) A drug can behave as a hapten and induce antibody production; in rare subjects, this happens, for instance, with penicillin. Upon a subsequent exposure, red cells are caught, as innocent bystanders, in the reaction between penicillin and antipenicillin antibodies. Hemolysis will subside as soon as penicillin administration is stopped. (2) A drug can trigger, perhaps through mimicry, the production of an antibody against a red cell antigen. The best-known

example is methyldopa, an antihypertensive agent no longer in use, which in a small fraction of patients stimulated the production of the Rhesus antibody anti-e. In patients who have this antigen, the anti-e is a true autoantibody, which then causes true AIHA (see above). Usually this will gradually subside once methyldopa is discontinued.

Severe intravascular hemolysis can be caused by the venom of certain snakes (cobras and vipers), and HA can also follow spider bites.

Paroxysmal Nocturnal Hemoglobinuria (PNH) PNH is an acquired chronic HA characterized by persistent intravascular hemolysis with occasional or frequent recurrent exacerbations. In addition to (i) hemolysis, there may be (ii) pancytopenia and (iii) a distinct tendency to venous thrombosis. This triad makes PNH a truly unique clinical condition; however, when not all of these three features are manifest on presentation, the diagnosis is often delayed, although it can always be made by appropriate laboratory investigations (see below).

PNH is encountered in all populations throughout the world, but it is a rare disease, with an estimated prevalence of ~5 per million (it may be somewhat less rare in Southeast Asia and in the Far East). PNH has about the same frequency in men and women. PNH is not inherited, and it has never been reported as a congenital disease, but it can present in small children or as late as in the seventies, although most patients are young adults.

CLINICAL FEATURES When seeking medical attention, the patient may report that one morning, she or he “passed blood instead of urine.” This distressing or frightening event may be regarded as the classic presentation; however, more frequently, this symptom is not noticed or not reported. Indeed, the patient often presents simply as a problem in the differential diagnosis of anemia, whether symptomatic or discovered incidentally. Sometimes the anemia is associated from the outset with neutropenia, thrombocytopenia, or both, thus signaling an element of bone marrow failure (see below). Some patients may present with recurrent attacks of severe abdominal pain eventually found to be related to thrombosis in abdominal veins, or attributable to NO depletion associated with intravascular hemolysis. When thrombosis affects the hepatic vein, it may produce acute hepatomegaly and ascites,

i.e., a full-fledged Budd-Chiari syndrome, which, in the absence of liver disease, ought to raise the suspicion of PNH.

The natural history of PNH can extend over decades. In the past, with supportive treatment only, the median survival was estimated to be about 10–20 years, with the most common cause of death being venous thrombosis, followed by infection secondary to severe neutropenia and hemorrhage secondary to severe thrombocytopenia. Rarely (estimated 1–2% of all cases), PNH may terminate in acute myeloid leukemia. On the other hand, full spontaneous recovery from PNH has been documented, albeit rarely.

LABORATORY INVESTIGATIONS AND DIAGNOSIS The most consistent blood finding is anemia, which may range from mild to moderate to very severe. The anemia is usually normo-macrocytic, with unremarkable red cell morphology. If the MCV is high, it is usually largely accounted for by reticulocytosis, which may be quite marked (up to 20%, or up to 400,000/ μ L). The anemia may become microcytic if the patient is allowed to become iron-deficient as a result of chronic iron loss through hemoglobinuria. Unconjugated bilirubin is mildly or moderately elevated; LDH is typically markedly elevated (values in the thousands are common); and haptoglobin is usually undetectable. All of these findings make the diagnosis of HA compelling. Hemoglobinuria may be overt in a random urine sample; if it is not, it may be helpful to obtain serial urine samples (Fig. 100-9) because hemoglobinuria can vary dramatically from day to day and even from hour to hour. The bone marrow is usually cellular, with marked to massive erythroid hyperplasia, often with mild to moderate dyserythropoietic features (these overlap with those seen in myelodysplastic syndromes, but PNH remains a separate entity). At some stage of the disease, the marrow may become hypocellular or even frankly aplastic (see below).

The definitive diagnosis of PNH must be based on the demonstration that a substantial proportion of the patient's red cells have an increased susceptibility to complement (C), due to the deficiency on their surface of proteins (particularly CD59 and CD55) that normally protect the red cells from activated C. The sucrose hemolysis test is unreliable; in contrast, the acidified serum (Ham) test is highly reliable but is carried out only in a few laboratories. The gold standard today is flow cytometry, which can be carried out on granulocytes as well as on red cells and has a very high sensitivity. In PNH, characteristically, one sees a bimodal distribution of cells, with a discrete population that is CD59 and CD55 negative. Although very small populations of CD59(−) cells are of interest in terms of pathophysiology (particularly of aplastic anemia [AA]), no patient should be diagnosed with PNH unless the proportion is substantial: in first approximation at least 5% of the total red cells and at least 20% of the total granulocytes.

PATHOPHYSIOLOGY Hemolysis in PNH is mainly intravascular and is due to an intrinsic abnormality of the red cell, which makes it exquisitely sensitive to activated C, whether C is activated through the alternative pathway or through an antigen-antibody reaction (classic pathway). The former mechanism is mainly responsible for chronic hemolysis in PNH; the latter explains why the hemolysis can be dramatically exacerbated in the course of a viral or bacterial infection. Hypersusceptibility to C is due to deficiency in the red cell membrane of several protective proteins (Fig. 100-10), among which CD59 is the most important because it is able to hinder the insertion into the membrane of C9 polymers (the so-called membrane attack complex, or MAC). The molecular basis for the deficiency of these proteins has been pinpointed not to a defect in any of the respective genes, but rather to the shortage of a unique glycolipid molecule, GPI (Fig. 100-2), which, through a peptide bond, anchors these proteins to the surface membrane of cells. The shortage of GPI is due in turn to a somatic mutation in an X-linked gene, called *PIGA*, required for an early step in GPI biosynthesis. As a result, the patient's marrow is a mosaic of mutant and nonmutant cells, and the peripheral blood always contains both GPI-negative (PNH) cells and GPI-positive (non-PNH) cells: in most cases the former prevail. Thrombosis is one of the most immediately life-threatening complications of PNH, and yet one of the least understood in its pathogenesis. It could be that deficiency of CD59 on the PNH platelet causes inappropriate platelet activation; however,

other mechanisms are possible. In very rare cases PNH can be caused by biallelic mutations of the *PIGT* gene, in the absence of a *PIGA* mutation. In these cases, because GPI is produced but cannot bind to proteins, the clinical picture is further complicated by the coexistence of a chronic inflammatory state.

BONE MARROW FAILURE (BMF) AND RELATIONSHIP BETWEEN PNH AND APLASTIC ANEMIA (AA) It is not unusual that patients with firmly established PNH have a previous history of AA, sometimes well documented; indeed, BMF preceding overt PNH is probably the rule rather than the exception. On the other hand, sometimes a patient with PNH becomes less hemolytic and more pancytopenic and ultimately has the clinical picture of AA. The relationship between PNH and AA manifested in the clinical course of patients may reflect a close link in pathogenesis. AA is thought to be an organ-specific autoimmune disease, in which T cells cause damage to hematopoietic stem cells via an as yet unidentified molecular target. The same may be true of PNH, and in this condition the target might be the GPI molecule itself. This would explain why GPI-negative (PNH) stem cells are spared; *PIGA* mutations can be demonstrated in normal people. Thus, PNH results from the combined action of two factors: failure of normal hematopoiesis and massive expansion of a PNH clone. There is evidence from mouse models that PNH stem cells do not expand on their own, and there is evidence from human patients that expansion is associated with negative selection against GPI-positive cells by GPI-specific T cells. Thus, PNH is a prime example of a clonal disease that is not malignant.

TREATMENT

Paroxysmal Nocturnal Hemoglobinuria

Until some 15 years ago there were essentially two treatment options for PNH: either allogeneic BMT, providing a definitive cure at the cost of nonnegligible risks; or continued supportive treatment for what, unlike other acquired HAs, may be a lifelong condition. A major advance has been the introduction in 2007 of a humanized monoclonal antibody, eculizumab, which binds to the complement component C5 near the site that, when cleaved, will trigger the distal part of the complement cascade leading to formation of the MAC. With C5 blocked by anti-C5, the patient is relieved of intravascular hemolysis and of its attendant consequences, including hemoglobinuria; with a substantial decrease in the rate of thrombosis. In the majority of those patients who needed regular blood transfusion, the transfusion requirement is either abolished or significantly reduced. For many PNH patients, eculizumab has meant a real improvement in the quality of life, as well as a decrease in complications, particularly thrombosis. At the same time, it is important to know that in patients on eculizumab the PNH red cells, now protected from being lysed through the MAC, do still bind C3 fragments and thus become opsonized. Therefore, hemolysis continues, but it is now extravascular. The extent to which this happens depends in part on a genetic polymorphism of the complement receptor CR1. Those patients who, on eculizumab, are still receiving blood transfusion are at risk of iron overload. Based on its half-life, eculizumab must be administered intravenously every 14 days. Ravulizumab, a long-lived anti-C5 derivative of eculizumab, is administered at 8-week rather than 2-week intervals: it provides similar benefit with obvious practical advantage.

Eculizumab and ravulizumab are very expensive and for this reason not accessible to patients in many parts of the world. Therefore, the management of PNH by supportive treatment is still very important. Folic acid supplements (at least 3 mg/d) are mandatory; the serum iron should be checked periodically, and iron supplements should be administered as appropriate. Transfusion of white cell-free red cells should be used whenever necessary, which, for some patients, means quite frequently. Long-term glucocorticoids are not indicated because there is no evidence that they have any effect on chronic hemolysis; in fact, they are contraindicated

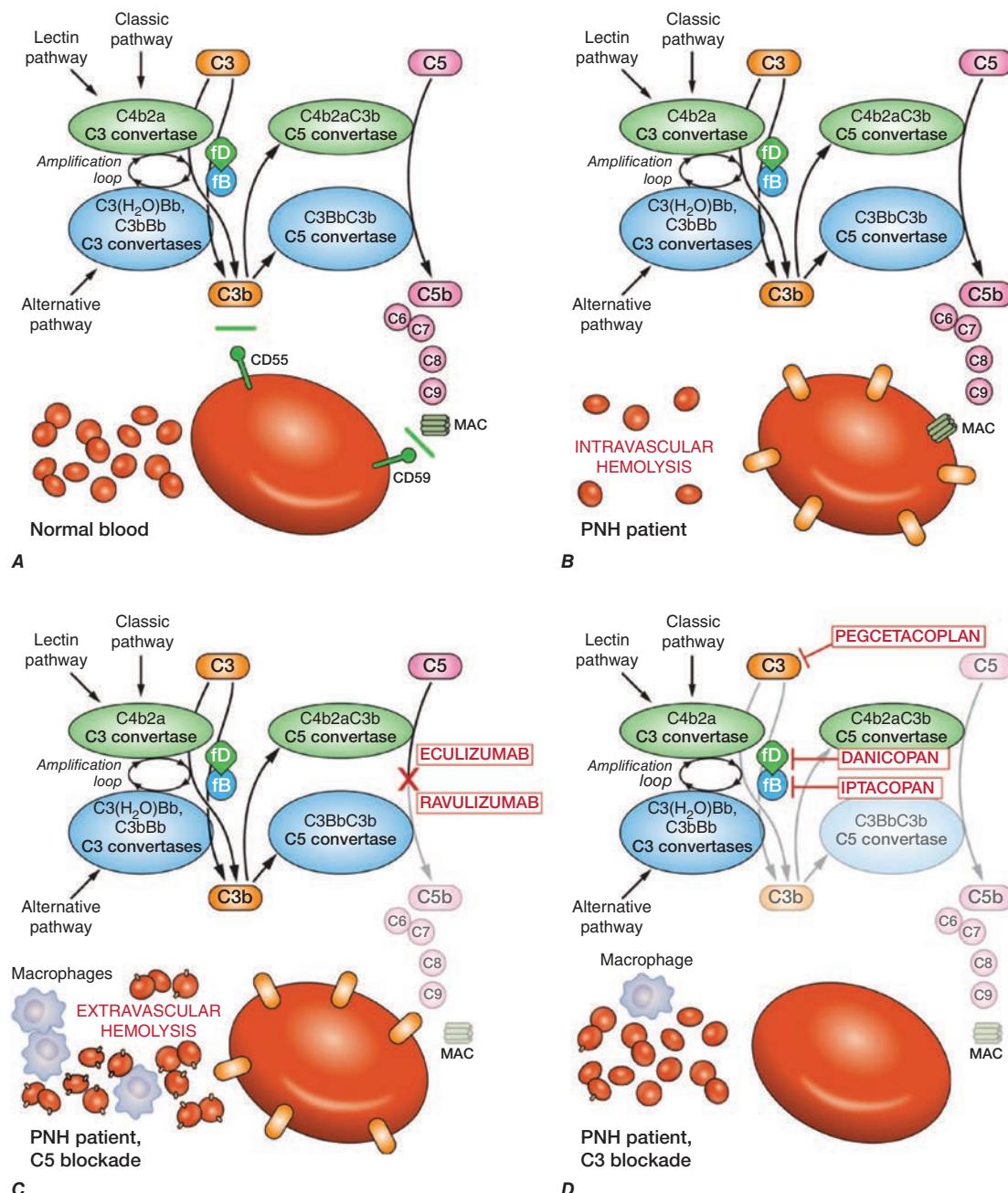


FIGURE 100-10 The complement cascade and the fate of red cells. **A**, In normal blood, when complement is activated, red cells are protected from lysis in several ways: primarily by the 2 glycosylphosphatidylinositol (GPI)-linked surface proteins CD55 (prevents binding of C3 fragments) and CD59 (prevents the membrane attack complex [MAC] from inserting into the membrane). **B**, PNH red cells are deficient in CD55 and CD59 because the GPI biosynthetic pathway is blocked as a result of a PIGA mutation; therefore, C3 fragments, particularly C3d, bind to their surface, and the red cells are rapidly lysed by the action of the MAC. **C**, With drugs (monoclonal antibodies) that bind to C5 and prevent it splitting into C5a and C5b, the entire distal pathway from C5 onward is blocked, MAC is not formed, and IVH is abrogated. However, red cells opsonized by C3d will be destroyed in the spleen and elsewhere; this drug-induced EVH varies in severity between patients. The Coombs test, which is characteristically negative in PNH, becomes positive (provided that a "broad spectrum" or an anticomplement reagent is used). **D**, With a drug that targets C3, C3b formation is inhibited, and the distal pathway is not triggered by C3b. Therefore, again, no MAC is formed (abrogating IVH), and, at the same time, opsonization of red cells by C3d is prevented, so that EVH is also curbed. The same is largely true for drugs that target factor B or factor D, although C3b can still be formed through the classical pathway. (Reproduced with permission from L Luzzatto: Control of hemolysis in patients with PNH. *Blood* 138:1909, 2021.)

because their side effects are considerable. A short course of prednisone may be useful when an inflammatory process exacerbates hemolysis. Any patient who has had venous thrombosis or who has a genetically determined thrombophilic state in addition to PNH should be on regular anticoagulant prophylaxis. With thrombotic complications that do not resolve otherwise, thrombolytic treatment with tissue plasminogen activator may be indicated.

Where anti-C5 therapy is available the proportion of PNH patients receiving BMT has decreased significantly. However, when an HLA-identical sibling is available, BMT should be taken into

consideration for any young patient with severe PNH; and for patients with the so-called PNH-AA syndrome, since eculizumab has no effect on BMF. For these patients immunosuppressive treatment with antithymocyte globulin and cyclosporine A may be an alternative, and it may be compatible with concurrent administration of eculizumab.

In view of persistent extravascular hemolysis, and sometimes persistent blood transfusion requirement in PNH patients on C5 blockade therapy, there has been great stimulus to developing agents that may inhibit complement activation more upstream. Several compounds that inhibit either the convertase function of C3 or plasma factors required for this function are currently in clinical trials (see Fig. 100-11).

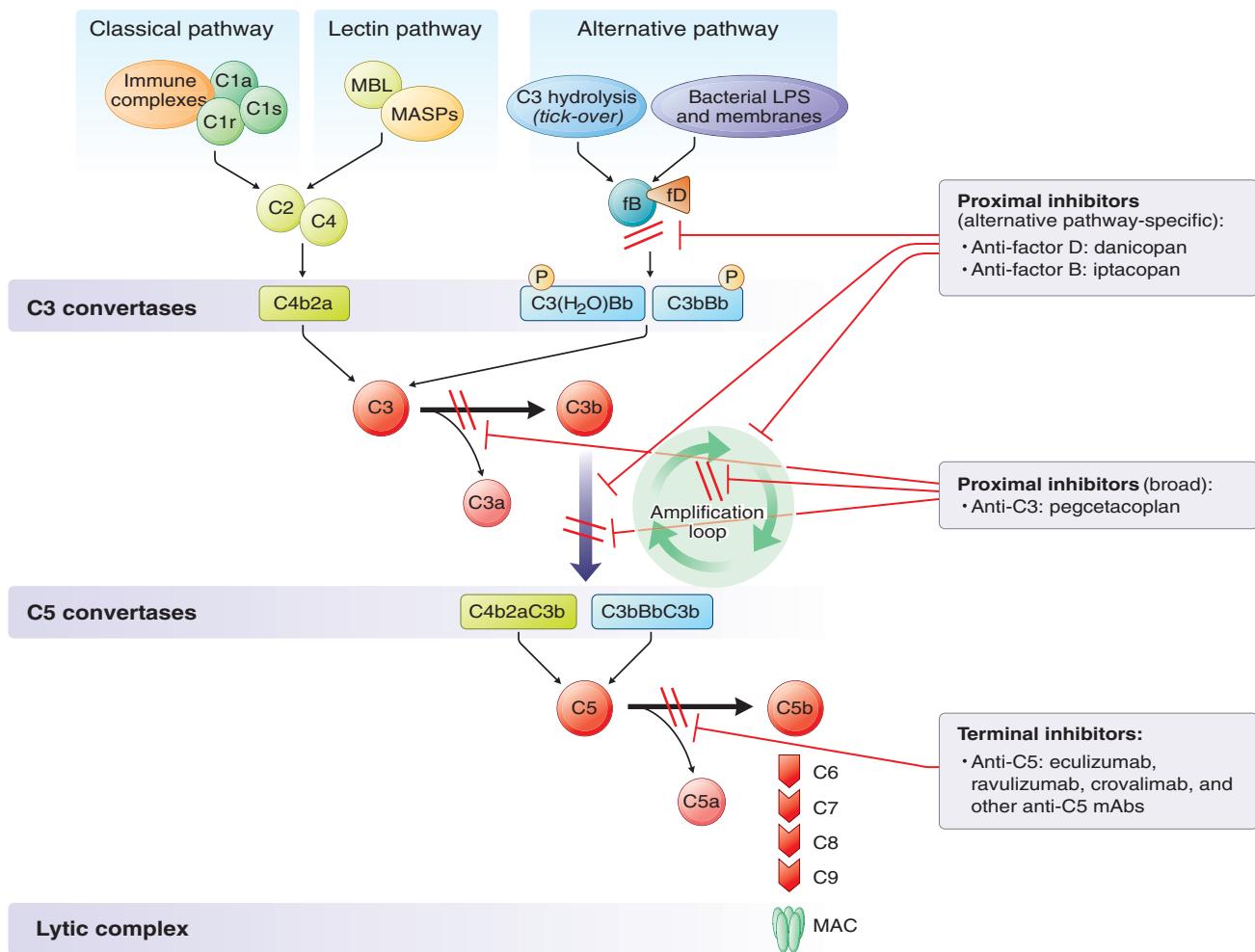


FIGURE 100-11 Monoclonal antibodies and small molecules in use or in development for the management of PNH and other complement-related disorders. Complement components are indicated by C followed by a number. MBL stands for mannose-binding lectin; MASPs for mannose-binding lectin-associated serine protease 1. P is properdin. Of the inhibitors shown on the right, only eculizumab and ravulizumab, which bind to C5 and are therefore inhibitors of the distal pathway, are already licensed drugs: both effectively abrogate MAC formation but they do not interfere with the formation of either the C3 convertase or the C5 convertase: in contrast, this can be achieved with the upstream inhibitors danicopan, iptacopan, and pegcetacoplan.

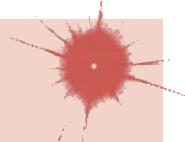
FURTHER READING

- BARCELLINI W et al: The changing landscape of autoimmune hemolytic anemia. *Front Immunol* 11:1, 2020.
- BRODSKY RA: Warm autoimmune hemolytic anemia. *N Engl J Med* 381:647-654, 2019.
- DACIE J: *The Haemolytic Anaemias*. London, Churchill Livingstone, volumes 1-5, 1985-1999.
- DE FRANCESCHI L et al: Acute hemolysis by hydroxychloroquine was observed in G6PD-deficient patient with severe COVID-19 related lung injury. *Eur J Intern Med* 77:136, 2020.
- GRACE RF et al: Clinical spectrum of pyruvate kinase deficiency: Data from the Pyruvate Kinase Deficiency Natural History Study. *Blood* 131:2183, 2018.
- IOLASCON A et al: Advances in understanding the pathogenesis of red cell membrane disorders. *Br J Haematol* 187:13, 2019.
- LOIRAT C et al: An international consensus approach to the management of atypical hemolytic uremic syndrome in children. *Pediatr Nephrol* 31:15, 2016.
- LUZZATTO L, KARADIMITRIS A: Paroxysmal nocturnal haemoglobinuria (PNH): Novel therapies for an ancient disease. *Br J Haematol* 191:579, 2020.
- LUZZATTO L et al: Glucose-6-phosphate dehydrogenase deficiency. *Blood* 136:1225, 2020.
- UYOGA S et al: Glucose-6-phosphate dehydrogenase deficiency and the risk of malaria and other diseases in children on the coast of Kenya: A case-control and a cohort study. *Lancet Haematology* 2:e437, 2015.

101

Anemia Due to Acute Blood Loss

Dan L. Longo



Blood loss causes anemia by two main mechanisms: (1) by the direct loss of red cells; and (2) if the loss of blood is protracted, it will gradually deplete iron stores, eventually resulting in iron deficiency. The latter type of anemia is covered in **Chap. 97**. Here, we are concerned with the former type, that is, *posthemorrhagic anemia*, which follows *acute* blood loss. This can be *external* (e.g., after trauma or obstetric hemorrhage) or *internal* (e.g., from bleeding in the gastrointestinal tract, rupture of the spleen, rupture of an ectopic pregnancy, subarachnoid hemorrhage, leaking aneurysm). In any of these cases, after the sudden loss of a large amount of blood, there are three clinical/pathophysiologic stages. (1) At first, the dominant feature is hypovolemia, which poses a threat particularly to organs that normally have a high blood supply, like the brain and the kidneys; therefore, loss of consciousness and acute renal failure are major threats. It is important to note that at this stage an ordinary blood count will not show anemia because the hemoglobin concentration is not affected. On physical exam, tachycardia, tachypnea, decreased pulse pressure, cold skin that appears pale and mottled, and decreased urine output may be noted. (2) Next, as

an emergency response, baroreceptors and stretch receptors will cause release of vasopressin and other peptides, and the body will shift fluid from the extravascular to the intravascular compartment, producing hemodilution; thus, the hypovolemia gradually converts to anemia. The degree of anemia will reflect the amount of blood lost. If after 3 days the hemoglobin is, for example, 7 g/dL, it means that about half of the entire blood has been lost. (3) Provided bleeding does not continue, the bone marrow response will gradually ameliorate the anemia. In this phase of the process, the reticulocyte count and erythropoietin levels will be elevated. The physiologic increase in marrow red cell production reflected by the increase in reticulocytes is similar to the marrow response to hemolysis.

The diagnosis of acute posthemorrhagic anemia (APHA) is usually straightforward, although sometimes internal bleeding episodes (e.g., after a traumatic injury), even when large, may not be immediately obvious. Look for physical findings that may help localize the bleeding. Grey Turner sign (flank ecchymosis) may reflect retroperitoneal bleeding. Cullen sign (umbilical ecchymosis) may suggest intraperitoneal or retroperitoneal bleeding. Dullness to chest percussion may suggest intrapleural bleeding. Whenever an abrupt fall in hemoglobin has taken place, whatever history is given by the patient, APHA should be suspected. Supplementary history may have to be obtained by asking the appropriate questions, and appropriate investigations (e.g., a sonogram or an endoscopy) may have to be carried out.

TREATMENT

Anemia Due to Acute Blood Loss

In patients who are hemodynamically unstable, the usual airway, breathing, and circulation assessments take priority. In the face of bleeding associated with hypotension, pharmacologic support with vasoressors is critical. With respect to anemia treatment, a two-pronged approach is imperative. (1) In many cases, the blood lost needs to be replaced promptly. Unlike with many chronic anemias, when finding and correcting the cause of the anemia is the first priority and blood transfusion may not be even necessary because the body is adapted to the anemia, with acute blood loss, the reverse is true; because the body is not adapted to the anemia, blood transfusion takes priority. (2) While the emergency is being confronted, it is imperative to stop the hemorrhage and to eliminate its source.

In an acute hemorrhage situation, plasma may be preferred to saline for volume expansion since dilution of clotting factors with crystalloid may interfere with hemostasis.

A special type of APHA is blood loss during and immediately after surgery, which can be substantial (e.g., up to 2 L in the case of a radical prostatectomy). Of course with elective surgical procedures, the patient's own stored blood may be available (through preoperative autologous blood donation), and in any case, blood loss ought to have been carefully monitored/measured. The fact that this blood loss is iatrogenic dictates that ever more effort should be invested in optimizing its management. The special features of transfusion medicine are discussed in [Chap. 113](#).

A Holy Grail of emergency medicine for a long time has been the idea of a blood substitute that would be universally available, suitable for all recipients, easy to store and to transport, safe, and as effective as blood itself. Two main paths have been pursued: (1) fluorocarbon synthetic chemicals that bind oxygen reversibly, and (2) artificially modified hemoglobins, known as hemoglobin-based oxygen carriers (HBOCs). Although there are numerous anecdotal reports of the use of both approaches in humans, and although HBOCs have reached the stage of phase 2–3 clinical trials, no “blood substitute” has yet become standard treatment.

FURTHER READING

HALDAR R et al: Artificial blood: A futuristic dimension of modern day transfusion sciences. *Cardiovasc Hematol Agents Med Chem* 17:11, 2019.

MARTINI WZ: Coagulation complications following trauma. *Mil Med Res* 3:35, 2016.

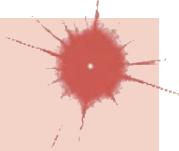
MORADI S et al: Artificial blood substitutes: First steps on the long route to clinical utility. *Clin Med Insights Blood Disord* 9:33, 2016.

MULLIER F et al: Facing coagulation disorders after acute trauma. *B-ENT Suppl* 26:67, 2016.

102

Bone Marrow Failure Syndromes Including Aplastic Anemia and Myelodysplasia

Neal S. Young



Bone marrow failure diseases include aplastic anemia, myelodysplastic syndrome (MDS), pure red cell aplasia (PRCA), and myelophthisis. Hypoproliferative anemia is a cardinal feature of these disorders, but more frequent is *pancytopenia*: anemia, leukopenia, and thrombocytopenia. Low blood counts in marrow failure result from deficient hematopoiesis, as distinguished from blood count depression due to peripheral destruction of red cells (hemolytic anemias), platelets (idiopathic thrombocytopenic purpura [ITP] or due to splenomegaly), and granulocytes (as in the immune leukopenias). Marrow damage and dysfunction also may be secondary to infection, inflammation, or cancer.

Hematopoietic failure syndromes are classified by dominant morphologic features of the bone marrow ([Table 102-1](#)). Although practical distinction among these syndromes usually is clear from the marrow pathology, some processes are so closely related that the diagnosis may be complex. Separation between aplastic anemia and hypocellular MDS can be particularly difficult. Mutations on genomic screens may be etiologic or interpreted as risk factors. Patients may seem to suffer from two or three related diseases simultaneously, or one diagnosis may appear to evolve into another. Many of these syndromes share an immune-mediated mechanism of marrow destruction and some element of genomic instability resulting in a higher rate of malignant transformation.

It is important that the internist and general practitioner recognize the marrow failure syndromes because quality of life and ultimate prognosis may be poor if the patient is untreated; effective therapies are often available but sufficiently complicated in their choice and delivery so as to warrant the care of a hematologist or oncologist. While the identification of pathogenic mutations on genomic screen, often on testing ordered by the internist and pediatrician, has revolutionized the diagnosis of the marrow failure syndromes, these results often require the interpretation of the hematologist and oncologist.

APLASTIC ANEMIA

■ DEFINITION

Aplastic anemia is pancytopenia with bone marrow hypocellularity. Acquired aplastic anemia is distinguished from iatrogenic aplasia, from marrow hypocellularity after intensive cytotoxic chemotherapy for cancer, and from usually accidental physical and chemical injury, as in radiation poisoning. Aplastic anemia can also be constitutional. Genetic diseases such as Fanconi anemia and dyskeratosis congenita usually (but not always) present in early childhood and have typical physical anomalies. Telomere diseases (see [Chap. 469](#)) and hematologic manifestations of mutations in genes such as GATA2, RUNX1, and MPL can present as marrow failure in normal-appearing adults.

TABLE 102-1 Differential Diagnosis of Pancytopenia**Pancytopenia with Hypocellular Bone Marrow**

Acquired aplastic anemia

Constitutional aplastic anemia (Fanconi anemia, dyskeratosis congenita, and others)

Hypocellular myelodysplastic syndrome

Rare aleukemic leukemia

Some acute lymphoid leukemia

Rare lymphomas of bone marrow

Copper deficiency

Pancytopenia with Cellular Bone Marrow

Primary bone marrow diseases

Myelodysplastic syndromes

Paroxysmal nocturnal hemoglobinuria (PNH)

Myelofibrosis

Aleukemic leukemia

Myelophthisis

Bone marrow lymphoma

Hairy cell leukemia

Secondary to systemic diseases

Systemic lupus erythematosus

Hypersplenism

 B_{12} folate deficiency

Copper deficiency

Alcohol

HIV infection

Brucellosis

Sarcoidosis

Tuberculosis

Leishmaniasis

Sepsis

Hypocellular Bone Marrow ± Pancytopenia

Q fever

Legionnaires' disease

Anorexia nervosa, starvation

Mycobacterium

Acquired aplastic anemia is often stereotypical in its manifestations, with the abrupt onset of low blood counts in a previously well young adult; seronegative hepatitis or a course of an incriminated medical drug may precede the onset. The diagnosis in these instances is uncomplicated. Sometimes blood count depression is moderate or incomplete, resulting in anemia, leukopenia, and thrombocytopenia in some combination. Aplastic anemia is related to both paroxysmal nocturnal hemoglobinuria (PNH; **Chap. 100**) and to MDS, and a clear distinction among these disorders may not be possible.

EPIDEMIOLOGY

The incidence of acquired aplastic anemia in Europe and Israel is two cases per million persons annually. In Thailand and China, rates of five to seven per million have been established. Men and women are affected with equal frequency, but the age distribution is biphasic, with the major peak in the teens and twenties and a second rise in older adults.

ETIOLOGY

The origins of aplastic anemia have been inferred from several recurring clinical associations (**Table 102-2**); unfortunately, these relationships are not reliable in an individual patient and may not be etiologic. In addition, although most cases of aplastic anemia are idiopathic, little other than history separates these cases from those with a presumed etiology such as a drug exposure.

Radiation Marrow aplasia is a major acute sequela of radiation. Radiation damages DNA; tissues dependent on active mitosis are particularly susceptible. Nuclear accidents involve not only power plant workers but also employees of hospitals, laboratories, and industry (food sterilization, metal radiography, etc.), as well as innocents exposed to stolen, misplaced, or misused sources. Whereas the radiation dose can be approximated from the rate and degree of decline in blood counts, dosimetry by reconstruction of the exposure can help to estimate the patient's prognosis and also to protect medical personnel

TABLE 102-2 Classification of Aplastic Anemia and Single Cytopenias

ACQUIRED	INHERITED/CONSTITUTIONAL
Aplastic Anemia	
Secondary	Fanconi anemia
Radiation	Dyskeratosis congenita/telomere disease
Drugs and chemicals	Shwachman-Diamond syndrome
Regular effects	Familial aplastic anemia/leukemia predisposition syndromes: <i>GATA2</i> , <i>RUNX1</i> , <i>CTLA4</i> , and others
Idiosyncratic reactions	
Viruses	Nonhematologic syndromes (Down, Dubowitz, Seckel)
Epstein-Barr virus (infectious mononucleosis)	
Hepatitis (non-A, non-B, non-C hepatitis)	
Parvovirus B19 (transient aplastic crisis, pure red cell aplasia [PRCA])	
HIV-1 (AIDS)	
Immune diseases	
Eosinophilic fasciitis	
Hypoimmunoglobulinemia	
Large granular lymphocytosis (LGL)	
Thymoma/thymic carcinoma	
Graft-versus-host disease in immunodeficiency	
Paroxysmal nocturnal hemoglobinuria (PNH)	
Pregnancy	
Idiopathic (immune)	
Cytopenias	
PRCA (see Table 102-4)	Congenital PRCA (Diamond-Blackfan anemia)
Neutropenia/agranulocytosis	Kostmann syndrome
Idiopathic	Shwachman-Diamond syndrome
Drugs, toxins	Reticular dysgenesis
LGL	
Pure white cell aplasia (+/- thymoma)	
Thrombocytopenia	
Drugs, toxins	Amegakaryocytic thrombocytopenia
Acquired amegakaryocytic thrombocytopenia	Thrombocytopenia with absent radii
	Other rare germline mutations

from contact with radioactive tissue and excreta. MDS and leukemia, but probably not aplastic anemia, are late effects of radiation.

Chemicals Benzene is a notorious cause of bone marrow failure: epidemiologic, clinical, and laboratory data link benzene to aplastic anemia, acute leukemia, and blood and marrow abnormalities. For leukemia, incidence is correlated with cumulative exposure, but susceptibility must also be important because only a minority of even heavily exposed workers develop myelotoxicity. The employment history is important, especially in industries where benzene is used for a secondary purpose, usually as a solvent. Benzene-related blood diseases have declined with regulation of industrial exposure. Although benzene is no longer generally available as a household solvent, exposure to its metabolites occurs in the normal diet and in the environment. The association between marrow failure and other chemicals is much less well substantiated. Further, there is scant direct evidence of marrow failure as a late effect of exposure, even to benzene.

Drugs (Table 102-3) Many chemotherapeutic drugs have marrow suppression as a major toxicity; effects are dose dependent and will

TABLE 102-3 Some Drugs and Chemicals Associated with Aplastic Anemia

Agents that regularly produce marrow depression as major toxicity in commonly used doses or normal exposures:
Cytotoxic drugs used in cancer chemotherapy: <i>alkylating agents, antimetabolites, antimitotics</i> , some antibiotics
Agents that frequently but not inevitably produce marrow aplasia:
Benzene
Agents associated with aplastic anemia but with a relatively low probability:
<i>Chloramphenicol</i>
Insecticides
Antiprotozoals: <i>quinacrine</i> and chloroquine, meprazine
Nonsteroidal anti-inflammatory drugs (including <i>phenylbutazone, indomethacin, ibuprofen, sulindac, aspirin</i>)
Anticonvulsants (<i>hydantoin, carbamazepine, phenacetin, felbamate</i>)
Heavy metals (<i>gold, arsenic, bismuth, mercury</i>)
Sulfonamides: some antibiotics, antithyroid drugs (methimazole, methylthiouracil, propylthiouracil), antidiabetes drugs (tolbutamide, chlorpropamide), carbonic anhydrase inhibitors (acetazolamide and methazolamide)
Antihistamines (<i>cimetidine, chlorpheniramine</i>)
D-Penicillamine
Estrogens (in pregnancy and in high doses in animals)
Agents whose association with aplastic anemia is more tenuous:
Other antibiotics (streptomycin, tetracycline, methicillin, mebendazole, trimethoprim/sulfamethoxazole, flucytosine)
Sedatives and tranquilizers (chlorpromazine, prochlorperazine, piperacetazine, chlordiazepoxide, meprobamate, methyprylon)
Allopurinol
Methyldopa
Quinidine
Lithium
Guanidine
Potassium perchlorate
Thiocyanate
Carbamazole

Note: Terms set in italics show the most consistent association with aplastic anemia.

occur in all recipients. In contrast, idiosyncratic reactions to a large and diverse group of drugs may lead to aplastic anemia without a clear dose-response relationship. A large international study in Europe in the 1980s quantitated drug relationships, especially for nonsteroidal analgesics, sulfonamides, thyrostatic drugs, some psychotropics, penicillamine, allopurinol, and gold. Association does not equal causation: a drug may have been used to treat the first symptoms of bone marrow failure (antibiotics for fever or a preceding viral illness) or provoked the first symptom of a preexisting disease (petechiae by nonsteroidal anti-inflammatory agents administered to the thrombocytopenic patient). In the context of total drug use, idiosyncratic reactions, although individually devastating, are rare events. Risk estimates are usually lower when determined in population-based studies. Furthermore, the low absolute risk is also made more obvious: even a 10- or 20-fold increase in risk translates, in a rare disease, to just a handful of drug-induced aplastic anemia cases among hundreds of thousands of exposed persons.

Infections Transient, mild blood count depression is frequent in the course of many viral and bacterial infections. Aplastic anemia can rarely follow infectious mononucleosis. Parvovirus B19 does not usually cause generalized bone marrow failure.

Immunologic Diseases Aplasia is a major consequence and the inevitable cause of death in *transfusion-associated graft-versus-host disease* (GVHD) that can occur after infusion of nonirradiated blood

products to an immunodeficient recipient. Aplastic anemia is strongly associated with the rare collagen vascular syndrome eosinophilic fascitis that is characterized by painful induration of subcutaneous tissues (Chap. 360). Thymoma and hypoimmunoglobulinemia are occasional associations with aplastic anemia. Pancytopenia with marrow hypoplasia can also occur in systemic lupus erythematosus (SLE).

Hepatitis Posthepatitis marrow failure accounts for ~5% of etiologies in most series. Patients are usually young men who have recovered from a bout of liver inflammation 1–2 months earlier; the subsequent pancytopenia is very severe. The hepatitis is seronegative (non-A, non-B, non-C); intensive laboratory efforts including deep sequencing have not disclosed an infectious agent, and the hepatitis is presumed to be immune-mediated. Fulminant liver failure in childhood can follow seronegative hepatitis, and marrow failure occurs at a high rate in these patients.

Pregnancy Aplastic anemia very rarely may occur and recur during pregnancy and resolve with delivery or with spontaneous or induced abortion.

Paroxysmal Nocturnal Hemoglobinuria An acquired mutation in the *PIG-A* gene in a hematopoietic stem cell is required for the development of PNH, but *PIG-A* mutations probably occur commonly in normal individuals. If the *PIG-A* mutant stem cell proliferates, the result is a clone of progeny deficient in glycosylphosphatidylinositol-linked cell surface membrane proteins (Chap. 100). Small clones of deficient cells can be detected by sensitive flow cytometry tests in one-half or more of patients with aplastic anemia at the time of presentation. Functional studies of bone marrow from PNH patients, even those with mainly hemolytic manifestations, show evidence of defective hematopoiesis. Patients with an initial clinical diagnosis of PNH, especially younger individuals, may later develop frank marrow aplasia and pancytopenia; patients with an initial diagnosis of aplastic anemia may suffer later from hemolytic PNH years after recovery of blood counts.

Constitutional Syndromes Fanconi anemia, an autosomal recessive disorder, manifests as congenital developmental anomalies, progressive pancytopenia, and an increased risk of malignancy. Chromosomes in Fanconi anemia are susceptible to DNA cross-linking agents, the basis for a diagnostic assay. Patients with Fanconi anemia typically have short stature, café au lait spots, and anomalies involving the thumb, radius, and genitourinary tract. At least 17 different genetic defects (all but one with an identified gene) have been defined; the most common, type A Fanconi anemia, is due to a mutation in *FANCA*. Most of the Fanconi anemia gene products form a protein complex that activates *FANCD2* by monoubiquitination to play a role in the cellular response to DNA damage and especially interstrand cross-linking.

Diamond-Blackfan anemia (see below) and Shwachman-Diamond syndrome are ribosomopathies, genetic defects in ribosome assembly that are tissue specific. In Shwachman-Diamond syndrome, presentation is early in life with neutropenia, pancreatic insufficiency, and malabsorption; most patients have compound heterozygous mutations in *SBDS*.

In the telomeropathies, inherited genetic defects alter telomere repair or one of the shelterin protein components of the telomere. The pediatric syndrome dyskeratosis congenita is characterized by the triad of mucous membrane leukoplakia, dystrophic nails, reticular hyperpigmentation, and early development of aplastic anemia (Chap. 469). Dyskeratosis congenita is due to mutations in genes of the telomere repair complex, which acts to maintain telomere length in replicating cells: the X-linked variety is due to mutations in the *DKC1* (*dyskerin*) gene; the more unusual autosomal dominant type is due to mutation in *TERC*, which encodes an RNA template. Rarely, mutations can also occur in genes such as *TNF2* that encode shelterin proteins, which bind telomere DNA.

Mutations in *TERC* and *TERT*, which encodes the catalytic reverse transcriptase telomerase, have subtle and milder effects on hematopoietic function, and presentation in adults is not unusual. It manifests as moderate aplastic anemia, which can be chronic and not progressive,

and isolated macrocytic anemia or thrombocytopenia. Physical anomalies are usually not present, but early hair graying is a clue to the diagnosis. A detailed personal and family history may disclose pulmonary fibrosis and hepatic cirrhosis. Variable penetrance means that *TERT* and *TERC* mutations represent risk factors for marrow failure, as family members with the same mutations may have normal or only slight hematologic abnormalities but more subtle evidence of (compensated) hematopoietic insufficiency. Measurement of telomere length of peripheral blood leukocytes is a commercially available functional test.

■ PATHOPHYSIOLOGY

Bone marrow failure results from severe damage to the hematopoietic cell compartment. In aplastic anemia, replacement of the bone marrow by fat is apparent in the morphology of the biopsy specimen (Fig. 102-1) and magnetic resonance imaging (MRI) of the spine. Cells bearing the CD34 antigen, a marker of early hematopoietic cells, are greatly diminished, and in functional studies, committed and primitive progenitor cells are virtually absent; *in vitro* assays have suggested that the stem cell pool is reduced to $\leq 1\%$ of normal in severe disease at the time of presentation.

Constitutional Genetic Syndromes

An intrinsic stem cell defect exists for the constitutional aplastic anemias: in a critical DNA repair pathway in Fanconi anemia, manifested in the laboratory as chromosome damage and cell death on exposure to certain chemical agents. In the telomeropathies, inability to repair telomeres or to protect chromosome ends is the result of mutations in genes of the telomerase complex or the shelterin proteins; telomere defects limit the cell's capacity to proliferate. Mutations in the *GATA* and *RUNX* genes affect signal transduction and transcriptional regulation in hematopoietic gene networks.

Chemical and Drug Injury Extrinsic damage to the marrow follows massive physical or chemical insults such as high doses of radiation and toxic chemicals. For the more common idiosyncratic reaction to modest doses of medical drugs, altered drug metabolism has been invoked as a mechanism. The metabolic pathways of many drugs and chemicals, especially if they are polar and have limited water solubility, involve enzymatic degradation to highly reactive electrophilic compounds; these intermediates are toxic because of their propensity to bind to cellular macromolecules. For example, derivative hydroquinones and quinolones are responsible for benzene-induced tissue injury. Excessive generation of toxic intermediates or failure to detoxify the intermediates may be genetically determined and apparent only on specific drug challenge; the complexity and specificity of the pathways imply multiple susceptibility loci and would provide an explanation for the rarity of idiosyncratic drug reactions.

Immune-Mediated Stem Cell Destruction The recovery of marrow function in some patients prepared for bone marrow transplantation with antilymphocyte globulin first suggested that aplastic anemia might be immune mediated. Laboratory data, including animal models, support an important role for the immune system in aplastic anemia. Blood and bone marrow cells of patients can suppress normal hematopoietic progenitor cell growth, and removal of T cells from aplastic anemia bone marrow improves hematopoiesis *in vitro*. Increased numbers of activated cytotoxic T-cell clones usually

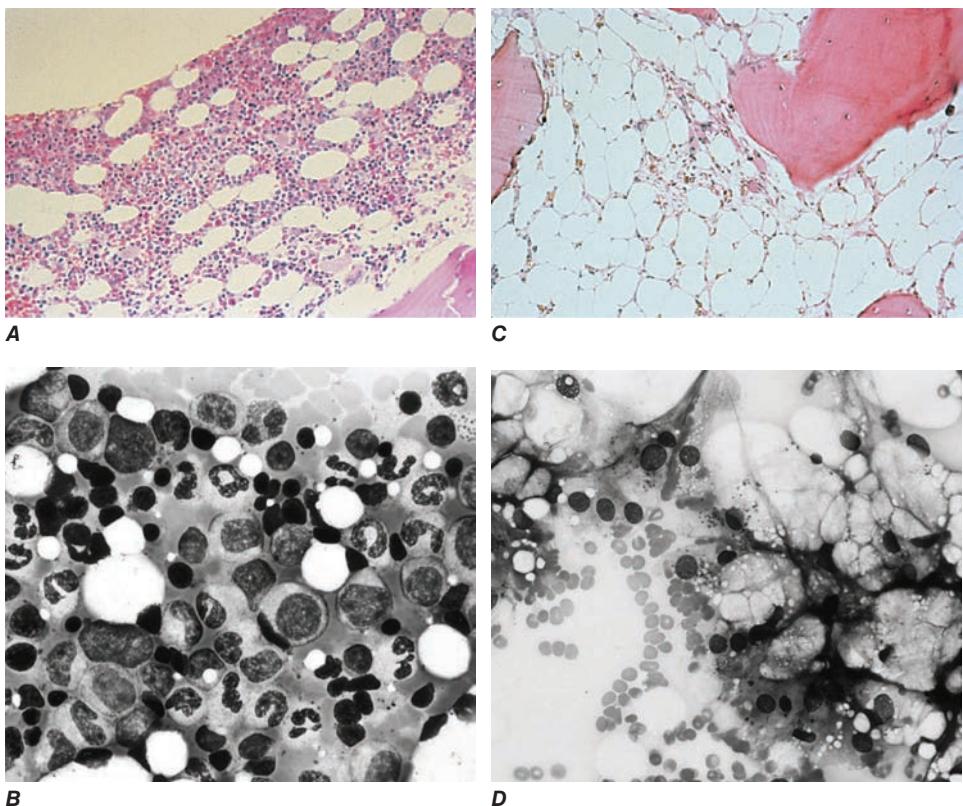


FIGURE 102-1 Normal and aplastic bone marrow. **A.** Normal bone marrow biopsy. **B.** Normal bone marrow aspirate smear. The marrow is normally 30–70% cellular, and there is a heterogeneous mix of myeloid, erythroid, and lymphoid cells. **C.** Aplastic anemia biopsy. **D.** Marrow smear in aplastic anemia. The marrow shows replacement of hematopoietic tissue by fat and only residual stromal and lymphoid cells.

decline with successful immunosuppressive therapy; type 1 cytokines are implicated; and interferon γ (IFN- γ) induces Fas expression on CD34 cells, leading to apoptotic cell death. Hematopoietic stem cells that have lost human leukocyte antigen (HLA) expression may be selectively expanded. The rarity of aplastic anemia despite common exposures (medicines, seronegative hepatitis) suggests that genetically determined features of the immune response can convert a normal physiologic response into a sustained abnormal autoimmune process, including polymorphisms in histocompatibility antigens, cytokine genes, and genes that regulate T-cell polarization (maturation toward helper or cytotoxic phenotypes) and effector function.

■ CLINICAL FEATURES

History Aplastic anemia can appear abruptly or insidiously. Bleeding is the most common early symptom; a complaint of days to weeks of easy bruising, oozing from the gums, nose bleeds, heavy menstrual flow, and sometimes petechiae will have been noticed. With thrombocytopenia, massive hemorrhage is unusual, but small amounts of bleeding in the central nervous system can result in catastrophic intracranial or retinal hemorrhage. Symptoms of anemia are also frequent, including lassitude, weakness, shortness of breath, and a pounding sensation in the ears. Infection is an unusual first symptom in aplastic anemia (unlike in agranulocytosis, where pharyngitis, anorectal infection, or frank sepsis occurs early). Patients often feel and look remarkably well despite drastically reduced blood counts. Systemic complaints and weight loss should point to other etiologies of pancytopenia. Prior medical drug use, chemical exposure, and preceding viral illnesses must often be elicited with directed questioning. A family history of hematologic diseases or blood abnormalities, of pulmonary or liver fibrosis, or of early hair graying points to a telomeropathy; a family history of unusual infections and warts points to *GATA2* deficiency.

Physical Examination Petechiae and ecchymoses are typical, and retinal hemorrhages may be present. Pelvic and rectal examinations

can often be deferred but, when performed, should be undertaken with great gentleness to avoid trauma; these may show bleeding from the cervical os and blood in the stool. Pallor of the skin and mucous membranes is common. Infection on presentation is unusual but may occur if the patient has been symptomatic for a few weeks. Lymphadenopathy and splenomegaly are highly atypical of aplastic anemia. Café au lait spots and short stature suggest Fanconi anemia; peculiar nails and leukoplakia suggest dyskeratosis congenita; early graying (and use of hair dyes to mask it!) suggests a telomerase defect.

■ LABORATORY STUDIES

Blood The smear shows large erythrocytes and a paucity of platelets and granulocytes. Mean corpuscular volume (MCV) is commonly increased. Reticulocytes are absent or few, and lymphocyte numbers may be normal or reduced. The presence of immature myeloid forms suggests leukemia or MDS; nucleated red blood cells (RBCs) suggest marrow fibrosis or tumor invasion; abnormal platelets suggest either peripheral destruction or MDS.

Bone Marrow The bone marrow is usually readily aspirated but dilute on smear, and the fatty biopsy specimen may be grossly pale on withdrawal; a “dry tap” instead suggests fibrosis or myelophthisis. In severe aplasia, the smear of the aspirated specimen shows only red cells, residual lymphocytes, and stromal cells; the biopsy (which should be >1 cm in length) is superior for determination of cellularity and shows mainly fat under the microscope, with hematopoietic cells occupying <25% of the marrow space; sometimes, the biopsy is virtually all fat. The correlation between marrow cellularity and disease severity is imperfect; patients with moderate disease by blood counts can have empty iliac crest biopsies, whereas “hot spots” of hematopoiesis may be seen in severe cases. Residual hematopoietic cells should have normal morphology, except for mildly megaloblastic erythropoiesis; megakaryocytes are greatly reduced and usually absent. Granulomas may indicate an infectious etiology of the marrow failure.

Ancillary Studies Chromosome breakage studies of peripheral blood using diepoxybutane or mitomycin C should be performed on children and younger adults to exclude Fanconi anemia. Very short telomere length strongly suggests the presence of a telomerase or shelterin mutation, which can be pursued by family studies and nucleotide sequencing. Chromosome studies of bone marrow cells are often revealing in MDS but should be negative in typical aplastic anemia. Flow cytometry offers a sensitive diagnostic test for PNH. Serologic studies may show evidence of recent viral infection, such as Epstein-Barr virus and HIV. Posthepatitis aplastic anemia is seronegative.

Genomics Next-generation sequencing allows for large number of genes to be tested for the presence of pathogenic mutations. Panels are available commercially and in certified academic laboratories. While expensive, they are very useful and sometimes critical in establishing the correct diagnosis. Germline gene panels examine 50 or more genes etiologic in constitutional bone marrow failure, including many for which functional assays (described above) are not available. A germline panel should be considered for all children and those adults with suggestive clinical features or family histories. Somatic mutations are sought when MDS is suspected. Myeloid neoplasm gene panels can query about 100 genes that are recurrently mutated in MDS and acute myeloid leukemia (AML). Pathogenic mutations in spliceosome genes and genes in the cohesion family are frequent in MDS and unexpected in aplastic anemia.

■ DIAGNOSIS

The diagnosis of aplastic anemia is usually straightforward, based on the combination of pancytopenia with a fatty bone marrow. Aplastic anemia is a disease of the young and should be a leading diagnosis in the pancytopenic adolescent or young adult. When pancytopenia is secondary, the primary diagnosis is usually obvious from either history or physical examination: the massive spleen of alcoholic cirrhosis, the history of metastatic cancer or SLE, or miliary tuberculosis on chest radiograph (Table 102-1).

Diagnostic problems can occur with atypical presentations and among related hematologic diseases. Patients with bone marrow hypocellularity may have depression of only one or two of three blood lines, with later progression to pancytopenia. The most important differential diagnoses are between acquired and constitutional aplastic anemia, and between aplastic anemia and MDS. The bone marrow in constitutional aplastic anemia is usually morphologically indistinguishable from the aspirate in acquired disease (an exception is GATA2 deficiency with its characteristic megakaryocyte atypia). The diagnosis can be suggested by family history, abnormal blood counts since childhood, or the presence of associated, sometimes subtle physical anomalies. Genomic testing for pathogenic mutations in genes etiologic in constitutional marrow failure syndromes can discriminate acquired from inherited aplastic anemia (but results may not return for several weeks, a problem in the severely pancytopenic patient). Acute myeloid leukemia (AML) and MDS in a pedigree should prompt screening for an inherited predisposition syndrome, such as RUNX1 mutations. Aplastic anemia may be difficult to distinguish from the hypocellular variety of MDS: MDS is favored by finding morphologic abnormalities, particularly of megakaryocytes and myeloid precursor cells, and typical cytogenetic abnormalities and somatic mutations on genomic screening of myeloid neoplasm genes (see above). There remains an unclear boundary between immune aplastic anemia and low-risk MDS: patients with deletion of 13q and 20q may respond well to immunosuppression, and mutations in genes such as DNMT3A and ASXL1 occur in both diseases.

■ PROGNOSIS

The natural history of severe aplastic anemia is rapid deterioration and death. Historically, provision first of RBCs and later of platelet transfusions and effective antibiotics were of some benefit, but few patients show spontaneous recovery. The major prognostic determinant is the blood count. Severe disease historically has been defined by the presence of two of three parameters: absolute neutrophil count <500/ μ L, platelet count <20,000/ μ L, and corrected reticulocyte count <1% (or absolute reticulocyte count <60,000/ μ L). In the era of effective immunosuppressive therapies, absolute numbers of reticulocytes (>25,000/ μ L) and lymphocytes (>1000/ μ L) may be better predictors of response to treatment and long-term outcome.

Other prognostic factors include the presence of a PNH clone, short telomeres on presentation, and somatically mutated white cells. Even small PNH clones may indicate an immune pathophysiology and responsiveness to immunosuppressive therapies. Telomere shortening in most patients likely reflects stem cell reserve, regenerative stress, and susceptibility to chromosomal instability. Collectively, the presence of mutations in the same myeloid neoplasia genes that are mutated in clonal hematopoiesis of indeterminate potential (ASXL1, DNMT3A) is associated with worse prognosis and clonal evolution.

TREATMENT

Aplastic Anemia

Severe acquired aplastic anemia can be cured by replacement of the absent hematopoietic cells (and the immune system) by stem cell transplant, or it can be ameliorated by suppression of the immune system to allow recovery of the patient’s residual bone marrow function. Glucocorticoids are not of value as primary therapy. Suspected exposures to drugs or chemicals should be discontinued; however, spontaneous recovery of severe blood count depression is rare, and a waiting period before beginning treatment may not be advisable unless the blood counts are only modestly depressed.

HEMATOPOIETIC STEM CELL TRANSPLANTATION

This is the first choice for the younger patient with a fully histocompatible sibling donor (Chap. 114). HLA typing should be ordered as soon as the diagnosis of aplastic anemia is established in a child or younger adult. In transplant candidates, transfusion of blood from family members should be avoided so as to prevent sensitization to histocompatibility antigens. In general, limited numbers of blood

products probably do not greatly affect outcome, especially when blood products are depleted of leukocytes. For allogeneic transplant from fully matched siblings, long-term survival rates for children are ~90%. Transplant morbidity and mortality are increased among adults, due to the higher risk of chronic GVHD and infections. Nevertheless, transplant should be considered early in all but the most elderly, including from alternative donors.

Most patients do not have a suitable sibling donor. Occasionally, a full phenotypic match can be found within the family and serve as well. Matched unrelated donors in large registries are available for the majority of Caucasian patients. With high-resolution matching at HLA, outcomes are similar to those with sibling donors, although complications (mainly GVHD and infection) are more frequent. Cord blood also can be a source of stem cells, especially for children. Matched unrelated donor transplants are often considered as initial treatment in children and as salvage therapy for adults after failed immunosuppression. Transplantation from an HLA haploidentical family donor is increasingly popular, as a donor is almost always quickly available. There is large experience in China, where lymphocyte depletion is usually performed before donor cell infusion. Posttransplant cyclophosphamide appears effective in preventing GVHD. Transplant protocols for marrow failure now usually do not include radiation in order to avoid late occurrence of cancer.

IMMUNOSUPPRESSION

The standard regimen of antithymocyte globulin (ATG) in combination with cyclosporine induces hematologic recovery (independence from transfusion and a leukocyte count adequate to prevent infection) in 60–70% of patients. Children do especially well, whereas older adult patients can suffer complications due to the presence of comorbidities. An early robust hematologic response correlates with long-term survival. Improvement in granulocyte number is generally apparent within 2 months of treatment. Most recovered patients continue to have some degree of blood count depression, the MCV remains elevated, and bone marrow cellularity returns toward normal very slowly if at all. Relapse (recurrent pancytopenia) is frequent, often occurring as cyclosporine is tapered or discontinued; most, but not all, patients respond to reinstitution of immunosuppression, but some responders become dependent on continued cyclosporine administration. “Clonal evolution,” isolated chromosomal abnormalities or the development of MDS, with typical cytogenetic aberrations and abnormal marrow morphology, occurs in ~15% of treated patients over a decade following initiation of ATG, usually but not invariably associated with a return of pancytopenia, and some patients develop leukemia. A laboratory diagnosis of PNH can generally be made at the time of presentation of aplastic anemia by flow cytometry; recovered patients may have frank hemolysis if the PNH clone expands. Bone marrow examinations should be performed if there is an unfavorable change in blood counts.

Horse ATG is administered as intravenous infusions and requires hospitalization. Rabbit ATG is much less effective, perhaps because it reduces T-regulatory cell numbers in patients. Serum sickness, a flulike illness with a characteristic cutaneous eruption and arthralgia, may develop ~10 days after initiating treatment. Methylprednisolone is administered with ATG to ameliorate the immune consequences of heterologous protein infusion. (Excessive or extended glucocorticoid therapy is associated with avascular joint necrosis.) Cyclosporine is administered orally at an initial high dose, with subsequent adjustment according to blood levels. Its most important side effects are nephrotoxicity, hypertension, and seizures.

Most patients with aplastic anemia lack a suitable marrow donor, and immunosuppression is the treatment of choice. Overall survival is equivalent with transplantation and immunosuppression. However, successful transplant cures marrow failure, whereas patients who recover adequate blood counts after immunosuppression remain at risk of relapse and malignant evolution. Increasing age

and the severity of neutropenia are the most important factors weighing in the decision between transplant and immunosuppression in adults who have a matched family donor: older patients do better with ATG and cyclosporine, whereas transplant is preferred if neutropenia is profound.

ELTROMBOPAG

Hematopoietic growth factors (HGFs) such as erythropoietin (EPO) and granulocyte colony-stimulating factor (G-CSF) are not effective in aplastic anemia, probably because endogenous blood levels in patients are extremely high. Circulating thrombopoietin is also elevated, but a thrombopoietin mimetic showed unexpected activity in refractory disease, producing robust, trilineage, and usually durable hematologic responses. Likely the mechanism of action of thrombopoietin mimetics is stimulation of the hematopoietic stem cell, but iron chelation and increased regulatory T cells are also possibly beneficial effects. Eltrombopag added to first-line immunosuppression with horse ATG markedly increased overall and complete response rates, to about 80% and 50%, respectively. Eltrombopag is approved by the U.S. Food and Drug Administration (FDA) as monotherapy for refractory aplastic anemia and in combination with horse ATG and cyclosporine as initial therapy.

Transplant from a suitable donor is preferred in the young patient, whereas immunosuppression is preferred in the older adult. Even heavily transfused and infected patients in whom immunosuppression has failed can be salvaged by stem cell transplant later.

ANDROGENS

The effectiveness of androgens has not been verified in controlled trials, but occasional patients will respond or even demonstrate blood count dependence on continued therapy. Sex hormones upregulate telomerase gene activity in vitro, which is possibly also their mechanism of action in improving marrow function. For patients with moderate disease, especially if a telomere gene defect is present, a 3- to 4-month trial may improve all blood counts (**Chap. 470**).

SUPPORTIVE CARE

Meticulous medical attention is required so that the patient may survive to benefit from definitive therapy or, having failed treatment, to maintain a reasonable existence in the face of pancytopenia. First and most important, infection in the presence of severe neutropenia must be aggressively treated by prompt institution of parenteral, broad-spectrum antibiotics. Therapy is empirical and must not await results of culture, although specific foci of infection such as oropharyngeal or anorectal abscesses, pneumonia, sinusitis, and typhlitis (necrotizing colitis) should be sought on physical examination and with radiographic studies. When indwelling plastic catheters become contaminated, vancomycin should be added. Persistent or recrudescent fever implies fungal disease: *Candida* and *Aspergillus* are common, especially after several courses of antibacterial antibiotics. A major reason for the improved prognosis in aplastic anemia has been the development of better antifungal drugs and the timely institution of such therapy when infection is suspected. Granulocyte transfusions can be effective when bacterial or fungal infection is progressive or refractory to antibiotics. Hand washing, the single best method of preventing the spread of infection, remains a neglected practice. Nonabsorbed antibiotics for gut decontamination are poorly tolerated and unproven, nor does reverse isolation reduce mortality from infections.

Both platelet and erythrocyte numbers can be maintained by transfusion. Alloimmunization historically limited the usefulness of platelet transfusions and is now minimized by several strategies, including use of single donors to reduce exposure and physical or chemical methods to diminish leukocytes in the product; HLA-matched platelets are usually effective in patients refractory to random donor products. Inhibitors of fibrinolysis such as aminocaproic acid have not been shown to relieve mucosal oozing; the use of low-dose glucocorticoids to induce “vascular stability” is unproven and not recommended. With prophylactic platelet transfusions, the

goal is to maintain the platelet count >10,000/ μL (oozing from the gut increases sharply at counts <5000/ μL). Menstruation should be suppressed either by oral estrogens or nasal follicle-stimulating hormone/luteinizing hormone antagonists. Aspirin and other nonsteroidal anti-inflammatory agents must be avoided in the presence of thrombocytopenia.

RBCs should be transfused so as to allow patient a normal level of activity, usually at a hemoglobin value of 70 g/L (90 g/L if there is underlying cardiac or pulmonary disease); a regimen of 2 units every 2 weeks will replace normal losses in a patient without a functioning bone marrow. In chronic anemia, the iron chelators deferoxamine and deferasirox should be added at approximately the fiftieth transfusion to avoid secondary hemochromatosis.

PURE RED CELL APLASIA

Other more restricted forms of marrow failure occur, in which only a single cell type is affected and the marrow shows corresponding absence or decreased numbers of specific precursor cells: aregenerative anemia as in PRCA (see below), thrombocytopenia with amegakaryocytosis (**Chap. 115**), and neutropenia without marrow myeloid cells in agranulocytosis (**Chap. 64**). In general, and in contrast to aplastic anemia and MDS, the unaffected lineages appear quantitatively and qualitatively normal. Agranulocytosis, the most frequent of these syndromes, is usually a complication of medical drug use, either by a mechanism of direct chemical toxicity or by immune destruction. Agranulocytosis has an incidence similar to aplastic anemia (but geographically more frequent in Europe than in Asia); in contrast to aplastic anemia, agranulocytosis is more prevalent among older adults and in women. Agranulocytosis should resolve with discontinuation of exposure, but significant mortality is attached to neutropenia in the older and often previously unwell patient. Both pure white cell aplasia (agranulocytosis without incriminating drug exposure) and amegakaryocytic thrombocytopenia are exceedingly rare and, like PRCA, appear to be due to a destructive immune response. In all of the single-lineage failure syndromes, progression to pancytopenia or leukemia is unusual.

DEFINITION AND DIFFERENTIAL DIAGNOSIS

PRCA is characterized by anemia, reticulocytopenia, and absent or rare erythroid precursor cells in the bone marrow. The classification of PRCA is shown in **Table 102-4**. In adults, PRCA is acquired. An identical syndrome can occur constitutionally: Diamond-Blackfan anemia, or congenital PRCA, is diagnosed at birth or in early childhood and often responds to glucocorticoid treatment; mutations in ribosome protein genes are etiologic. Temporary red cell failure occurs in transient aplastic crisis of hemolytic anemias due to acute parvovirus infection (**Chap. 197**) and in transient erythroblastopenia of childhood, which occurs in normal children.

CLINICAL ASSOCIATIONS AND ETIOLOGY

PRCA has important associations with immune system diseases. A minority of cases occur with a thymoma. More frequently, red cell aplasia can be the major manifestation of large granular lymphocytosis or complicates chronic lymphocytic leukemia. Some patients may be hypogammaglobulinemic. A ribosomal protein gene is deleted in the 5q- syndrome, such that the MDS may manifest as an acquired red cell aplasia. Occasionally (as compared to agranulocytosis), PRCA can be due to an idiosyncratic drug reaction. Subcutaneous administration of EPO has provoked PRCA mediated by neutralizing antibodies to the hormone. PRCA due to antibodies to blood group antigens (isoagglutins) is a complication of allogeneic stem cell transplant. For most PRCAs, T-cell inhibition is probably the prevalent immune mechanism.

PERSISTENT PARVOVIRUS B19 INFECTION

Chronic parvovirus infection is a treatable cause of red cell aplasia. This common virus causes a benign exanthem of childhood (fifth disease) and a polyarthralgia/arthritis syndrome in adults. In patients

TABLE 102-4 Classification of Pure Red Cell Aplasia

Self-limited	
	Transient erythroblastopenia of childhood
	Transient aplastic crisis of hemolysis (acute B19 parvovirus infection)
Fetal red blood cell aplasia	
	Nonimmune hydrops fetalis (in utero B19 parvovirus infection)
Constitutional pure red cell aplasia	
	Congenital pure red cell aplasia (Diamond-Blackfan anemia)
Acquired pure red cell aplasia	
	MDS (5q- syndrome)
	Cancer
	Thymoma
	Lymphoid malignancies (and more rarely other hematologic diseases)
	Paraneoplastic to solid tumors
Connective tissue disorders with immunologic abnormalities	
	Systemic lupus erythematosus, juvenile rheumatoid arthritis, rheumatoid arthritis
	Multiple endocrine gland insufficiency
Viruses	
	Persistent B19 parvovirus, hepatitis, adult T-cell leukemia virus, Epstein-Barr virus
Pregnancy	
Drugs	
	Especially phenytoin, azathioprine, chloramphenicol, procainamide, isoniazid
Antibodies to erythropoietin	
Idiopathic (immune)	

with underlying hemolysis (or any condition that increases demand for RBC production), parvovirus infection can cause a transient aplastic crisis and an abrupt but temporary worsening of the anemia due to failed erythropoiesis. In normal individuals, acute infection is resolved by production of neutralizing antibodies to the virus, but in the setting of congenital, acquired, or iatrogenic immunodeficiency, persistent viral infection may occur. The bone marrow shows red cell aplasia and the presence of giant pronormoblasts (**Fig. 102-2**), which is the cytopathic sign of B19 parvovirus infection. Viral tropism for human erythroid progenitor cells is due to its use of erythrocyte P antigen as a cellular receptor for entry. Direct cytotoxicity of virus causes anemia if demands on erythrocyte production are high; in normal individuals, the temporary cessation of red cell production is not clinically apparent, and skin and joint symptoms are mediated by immune complex deposition.

TREATMENT

Pure Red Cell Aplasia

History, physical examination, and routine laboratory studies may disclose an underlying disease or a drug exposure. Thymoma should be sought by radiographic procedures; tumor excision is indicated, but anemia does not necessarily improve with surgery. The diagnosis of parvovirus infection requires detection of viral DNA sequences in the blood (IgG and IgM antibodies are commonly absent). The presence of erythroid colonies has been considered predictive of response to immunosuppressive therapy in idiopathic PRCA.

Red cell aplasia is compatible with long-term survival with supportive care alone: a combination of erythrocyte transfusions and iron chelation. For persistent B19 parvovirus infection, almost all patients respond to intravenous immunoglobulin therapy. The majority of patients with acquired PRCA respond favorably to immunosuppression: glucocorticoids, cyclosporine, ATG, azathioprine, and cyclophosphamide are effective.

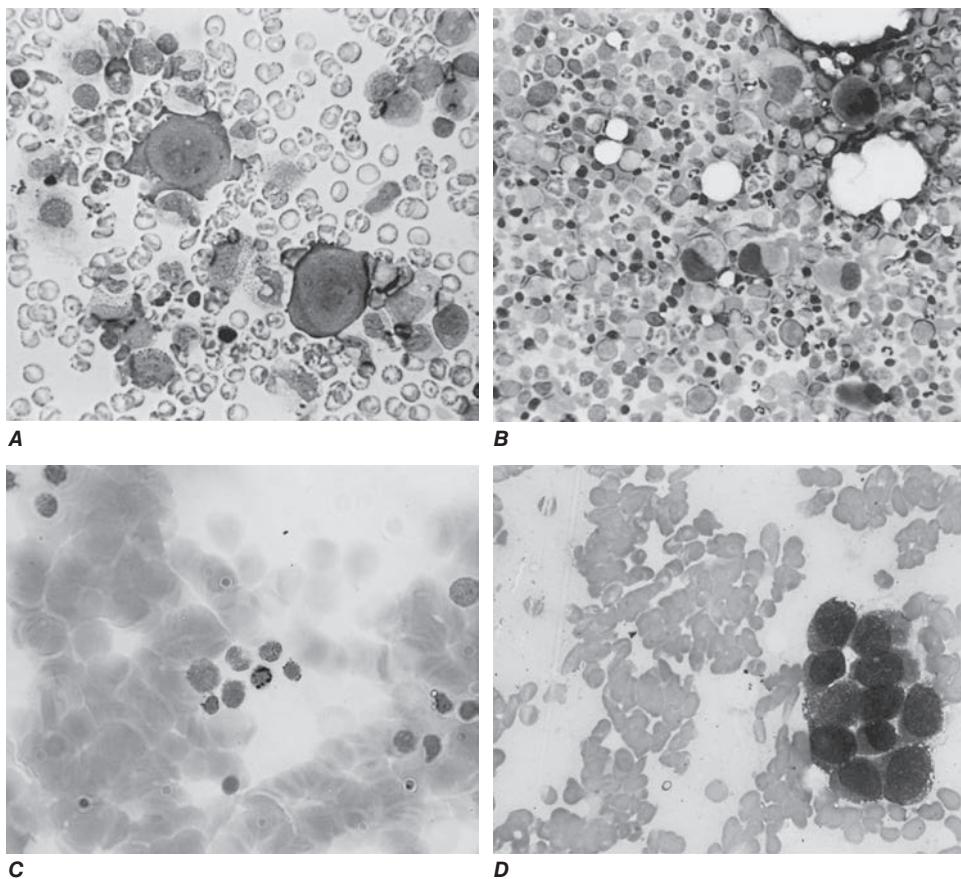


FIGURE 102-2 Pathognomonic cells in marrow failure syndromes. **A.** Giant pronormoblast, the cytopathic effect of B19 parvovirus infection of the erythroid progenitor cell. **B.** Uninuclear megakaryocyte and microblastic erythroid precursors typical of the 5q- myelodysplasia syndrome. **C.** Ringed sideroblast showing perinuclear iron granules. **D.** Tumor cells present on a touch preparation made from the marrow biopsy of a patient with metastatic carcinoma.

MYELODYSPLASTIC SYNDROMES

■ DEFINITION

The MDS are a heterogeneous group of hematologic disorders characterized by both (1) cytopenias due to bone marrow failure and (2) a high risk of development of AML. Anemia due to ineffective erythropoiesis, often with thrombocytopenia and neutropenia, occurs with dysmorphic (abnormal appearing) and usually cellular bone marrow, or with specific chromosome abnormalities or acquired mutations. In patients with “low-risk” MDS, marrow failure dominates the clinical course. In other patients, myeloblasts are present at diagnosis, chromosomes are abnormal, and the “high risk” is due to leukemic progression. MDS may be fatal due, most often, to complications of pancytopenia or to progression to leukemia, but a large proportion of patients will die of concurrent disease, the comorbidities typical in an elderly population. A useful nosology of these often-confusing entities was first developed by the French-American-British Cooperative Group in 1983. Five subtypes were defined then: refractory anemia (RA), refractory anemia with ringed sideroblasts (RARS), refractory anemia with excess blasts (RAEB), refractory anemia with excess blasts in transformation (RAEB-t), and chronic myelomonocytic leukemia (CMML). The World Health Organization (WHO) classification (2002) recognized that the distinction between RAEB-t and AML was arbitrary, grouped them together as acute leukemia, and clarified that CMML behaves as a myeloproliferative disease. The current WHO classification of 2016 is more refined but also more complicated (**Table 102-5**): blast percentage remains critical in defining MDS categories; erythroid predominant leukemias are now largely regarded as MDS; defining cytogenetic abnormalities are reaffirmed; and a single somatic mutation, in *SF3B1*, is now a feature of sideroblastic anemias. Identification of somatically mutated genes and their correlation with

clinical outcomes will be increasingly important in defining classification, prognosis, and targeting therapy.

The diagnosis of MDS can be a challenge, even for the expert, because sometimes subtle clinical and pathologic features must be distinguished, and precise diagnostic categorization requires a hematopathologist knowledgeable in the latest classification scheme. Unfortunately, agreement among pathologists on morphologic features and classification is imperfect; changes in the appearance of megakaryocytes are more reliable than loss of granules in neutrophil precursors or dyserythropoiesis. Further, dysplastic changes can be observed in normal individuals, and they can occur with vitamin deficiencies and as drug effects. Genomic testing is increasingly routine and can be difficult to interpret, as in differences between somatic and germline mutations, pathogenic mutations versus those of unknown significance (clonal hematopoiesis increases in frequency with age and involves genetic changes that may be clinically silent or convey an increased risk of hematologic malignancy), and clone size and changes over time. It is important that the internist and primary care physician be sufficiently familiar with MDS to expedite referral to a hematologist because many new therapies are now available to improve hematopoietic function and the judicious use of supportive care can improve the patient's quality of life.

■ EPIDEMIOLOGY

MDS is a disease of the elderly; the mean age at onset is older than 70 years. There is a slight male predominance. MDS is a relatively common form of bone marrow failure, with reported incidence rates of 35 to >100 per million persons in the general population and 120 to >500 per million in older adults. Estimates of incidence in the United States range from 30,000 to 40,000 new cases annually and a prevalence of 60,000–120,000 in the population. Rates of MDS have increased over time due to better recognition of the syndrome by physicians and an aging population.

MDS is rare in children, in whom it often has a constitutional genetic basis that can be identified on genomic screens of myeloid cancer predisposition panels.

Secondary or therapy-related MDS, usually related to previous iatrogenic exposure to alkylating agents and other chemotherapy as well as radiation, is not age related.

■ ETIOLOGY AND PATHOPHYSIOLOGY

MDS is associated with environmental exposures such as radiation and benzene; other risk factors have been reported inconsistently. Secondary, therapy-related MDS occurs as a late toxicity of cancer treatment; radiation and the radiomimetic alkylating agents such as busulfan, nitrosourea, or procarbazine (with a latent period of 5–7 years); or the DNA topoisomerase inhibitors (2-year latency). Acquired aplastic anemia, Fanconi anemia, and other constitutional marrow failure diseases can evolve into MDS; occasionally, MDS in adults is recognized as due to germline *GATA2*, *RUNX1*, or telomere gene mutations. The typical MDS patient does not have a suggestive environmental exposure history or a preceding hematologic disease. MDS is a disease of aging, consistent with accumulation of mutations within a hematopoietic stem cell in an aging marrow environment.

TABLE 102-5 World Health Organization (WHO) Classification of Myelodysplastic Syndromes (MDS)/Neoplasms

NAME	RING SIDEROBLASTS	MYELOBLASTS	KARYOTYPE
MDS with single lineage dysplasia (MDS-SLD)	<15% (<5%) ^a	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with multilineage dysplasia (MDS-MLD)	<15% (<5%) ^a	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with ring sideroblasts (MDS-RS)			
MDS-RS with single lineage dysplasia (MDS-RS-SLD)	≥15% / ≥5% ^a	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS-RS with multilineage dysplasia (MDS-RS-MLD)	≥15% / ≥5% ^a	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with isolated del(5q)	None or any	BM <5%, PB <1%, no Auer rods	del(5q) alone or with 1 additional abnormality except -7 or del(7q)
MDS with excess blasts (MDS-EB)			
MDS-EB-1	None or any	BM 5–9% or PB 2–4%, no Auer rods	Any
MDS-EB-2	None or any	BM 10–19% or PB 5–19% or Auer rods	Any
MDS, unclassifiable (MDS-U)			
• with 1% blood blasts	None or any	BM <5%, PB = 1%, no Auer rods	Any
• with single lineage dysplasia and pancytopenia	None or any	BM <5%, PB = 1%, no Auer rods	Any
• based on defining cytogenetic abnormality	<15%	BM <5%, PB = 1%, no Auer rods	MDS-defining abnormality
Refractory cytopenia of childhood	None	BM <5%, PB <2%	Any

^aIf *SF3B1* mutation is present.

Abbreviations: BM, bone marrow; PB, peripheral blood.

MDS is a clonal hematopoietic stem cell disorder characterized by disordered cell proliferation, impaired differentiation, and aberrant hematopoiesis, resulting in cytopenias and risk of progression to leukemia. Both chromosomal and genetic instability have been implicated; both are aging-related. Cytogenetic abnormalities are found in approximately one-half of patients, and some of the same specific lesions are also seen in leukemia; aneuploidy (chromosome loss or gain) is more frequent than translocations. Accelerated telomere attrition may destabilize the genome in marrow failure and predispose to acquisition of chromosomal lesions. Cytogenetic abnormalities are not random (loss of all or part of 5, 7, and 20, trisomy of 8) and may be related to etiology (11q23 following topoisomerase II inhibitors). The type and number of cytogenetic abnormalities strongly correlate with the probability of leukemic transformation and survival.

Genomics has illuminated the role of specific mutations and distinct molecular pathways in the pathophysiology of MDS. Somatic mutations in about 100 genes, which are recurrently present in myeloid neoplasms and are acquired in about 100 genes, are arise in the abnormal marrow cells (and are absent in the germline). Many of the same genes are mutated in AML and in MDS, whereas others are distinctive in subtypes of MDS. A prominent example is *SF3B1*, in which mutations strongly associate with sideroblastic anemia. Some mutations correlate with prognosis: spliceosome defects (like *SF3B1*) correlate with favorable outcome, and mutations in *EZH2*, *TP53*, *RUNX1*, and *ASXL1* with poor outcome. Correlation and exclusion in the pattern of mutations indicate a functional genomic architecture. Driver genes mutated early are consistent with normal blood counts and marrow morphology, but these expanded clones of cells containing them are susceptible to malignant transformation with the acquisition of additional mutations. Deep sequencing results in patients whose MDS evolved to AML have shown clonal succession, with founder clones acquiring additional mutations to produce clonal dominance. Mutations and cytogenetic abnormalities are not independent: *TP53* mutations associate with complex cytogenetic abnormalities and *TET2* mutations with normal cytogenetics. The prevalence of abnormal cells by morphology underestimates bone marrow involvement by MDS clones, as cells normal in appearance are derived from the abnormal clones. Presenting and evolving hematologic manifestations result from the accumulation of multiple genetic lesions: loss of tumor-suppressor genes, activating oncogene, epigenetic pathways that affect mRNA processing and methylation status, or other harmful alterations. Pathophysiology has been linked to mutations and chromosome abnormalities in some specific

MDS syndromes. The 5q- deletion leads to heterozygous loss of a ribosomal protein gene which mimics constitutional red cell aplasia. An immune pathophysiology may be important in lower risk MDS, as cytopenias can respond to immunosuppressive therapy as administered for aplastic anemia. In general for MDS, the role of the immune system and its cells and cytokines; the role of the hematopoietic stem cell niche, the microenvironment, and cell-cell interactions; the fate of normal cells in the Darwinian competitive environment of the dysplastic marrow; and how mutant cells produce marrow failure in MDS are still not completely understood.

CLINICAL FEATURES

Anemia dominates the early course. Most symptomatic patients complain of the gradual onset of fatigue and weakness, dyspnea, and pallor, but at least one-half of patients are asymptomatic, and their MDS is discovered only incidentally on routine blood counts. Previous chemotherapy or radiation exposure is an important historic fact. Fever and weight loss are more often features of a myeloproliferative rather than myelodysplastic process. MDS in childhood is rare and, when diagnosed, implicates an underlying genetic disease. Children with Down syndrome are susceptible to MDS as well as leukemia. A family history may indicate a hereditary form of sideroblastic anemia, Fanconi anemia, or a telomeropathy. Inherited *GATA2* mutations, as in the MonoMAC syndrome (with increased susceptibility to viral, mycobacterial, and fungal infections, as well as deficient numbers of monocytes, natural killer cells, and B lymphocytes), predispose to MDS. Germline *RUNX1* mutations also confer a high risk of MDS and leukemia, often preceded by years of modest thrombocytopenia. A family history is important in all MDS patients, as constitutional mutations may not result in manifest disease until adulthood.

The physical examination in MDS is remarkable for signs of anemia; approximately 20% of patients have splenomegaly. Some unusual skin lesions, including Sweet's syndrome (febrile neutrophilic dermatosis), occur with MDS. Accompanying autoimmune syndromes are not infrequent. In the younger patient, stereotypical anomalies point to a constitutional syndrome (short stature, abnormal thumbs in Fanconi anemia; early graying in the telomeropathies; cutaneous warts in *GATA2* deficiency).

LABORATORY STUDIES

Blood Anemia is present in most cases, either alone or as part of bi- or pancytopenia; isolated neutropenia or thrombocytopenia is

more unusual. Macrocytosis is common, as in most marrow failure disease. Platelets also are large and lack granules. In functional studies, they may show marked abnormalities, and patients may have bleeding symptoms despite seemingly adequate numbers. Neutrophils are hypo-granulated; have hyposegmented, ringed, or abnormally segmented nuclei; contain Döhle bodies; and may be functionally deficient. Circulating myeloblasts usually correlate with marrow blast numbers, and their quantity is important for classification and prognosis. The total white blood cell count (WBC) is usually normal or low, except in CMML. As in aplastic anemia, MDS can be associated with a clonal population of PNH cells. Genetic testing is commercially available for constitutional syndromes.

Bone Marrow The bone marrow is usually normal or hypercellular, but in about 20% of cases, it is sufficiently hypocellular to lead to confusion with aplastic anemia. No single characteristic feature of marrow morphology distinguishes MDS, but the following are commonly observed: dyserythropoietic changes (especially nuclear abnormalities) and ringed sideroblasts in the erythroid lineage; hypogranulation and hyposegmentation in granulocytic precursors, with an increase in myeloblasts; and megakaryocytes showing reduced numbers or disorganized nuclei. Megaloblastic nuclei and defective hemoglobinization in the erythroid lineage are common. Prognosis strongly correlates with the proportion of marrow blasts, which should be enumerated manually on the marrow smear and by flow cytometry of an aspirate. Flow cytometry can also reveal characteristically aberrant hematopoietic differentiation. Cytogenetics and fluorescent *in situ* hybridization can identify chromosomal abnormalities.

■ DIFFERENTIAL DIAGNOSIS

Deficiencies of vitamin B₁₂ or folate should be excluded by appropriate blood tests; vitamin B₆ deficiency can be assessed by a therapeutic trial of pyridoxine if the bone marrow shows ringed sideroblasts. Copper deficiency can lead to cytopenias and dysplastic marrows of varying cellularity. Marrow dysplasia can be observed in acute viral infections, drug reactions, or chemical toxicity but should be transient. More difficult are the distinctions between hypocellular MDS and aplasia or between RA with excess blasts and acute leukemia: the WHO considers 20% blasts in the marrow as the criterion that separates AML from MDS. In young patients, underlying, predisposing genetic diseases should be considered and appropriate genomic testing performed (see above).

■ PROGNOSIS

The median survival varies greatly from years for patients with 5q- or sideroblastic anemia to a few months in RA with excess blasts or severe pancytopenia associated with monosomy 7. The International Prognostic Scoring System (IPSS), revised in 2012 (Table 102-6), assists in making predictions. Even “lower-risk” MDS has significant morbidity and mortality. More refined (and also more complicated) prognostic

scoring systems can separate those with intermediate-1 risk who have relatively poor prognoses. Prognostic systems have been developed based on survival from diagnosis, but prognosis changes over time, and hazard ratios for survival and leukemic transformation converge over time among risk categories, consistent with dynamic changes in clonal architecture.

Most patients die as a result of complications of pancytopenia and not due to leukemic transformation; perhaps one-third succumb to diseases unrelated to their MDS. Precipitous worsening of pancytopenia, acquisition of new chromosomal abnormalities on serial cytogenetic determination, increase in the number of blasts, and marrow fibrosis are all poor prognostic indicators. The outlook in therapy-related MDS, regardless of type, is extremely poor, and most patients progress within a few months to refractory AML.

TREATMENT

Myelodysplasia

Historically, therapy of MDS has been unsatisfactory, but several drugs may not only improve blood counts but also delay onset of leukemia and improve survival. The choice of therapy for an individual patient, administration of treatment, and management of toxicities are complicated and require hematologic expertise.

Only hematopoietic stem cell transplantation offers cure of MDS. The survival rate in selected patient cohorts is ~50% at 3 years but improving. Results using unrelated matched donors are similar to those with siblings, and patients in their fifties and older have been successfully transplanted. Nevertheless, treatment-related mortality and morbidity increase with recipient age. The transplant conundrum is that the high-risk patient (by IPSS score and presence of monosomal karyotype), for whom the procedure is most obviously indicated, has a high probability of a poor outcome from transplant-related mortality or disease relapse, whereas the low-risk patient, who is more likely to tolerate transplant, also may do well for years with less aggressive therapies. In practice, only a small proportion of MDS patients undergo transplantation.

MDS has been regarded as particularly refractory to cytotoxic chemotherapy regimens, and as in AML in the older adult, drug toxicity is frequent and often fatal, and remissions, if achieved, are brief. Low doses of cytotoxic drugs have been administered for their “differentiation” potential, and from this experience, drug therapies have emerged based on pyrimidine analogues. These drugs are classified as epigenetic modulators, believed to act through a demethylating mechanism to alter gene regulation and allow differentiation to mature blood cells from the abnormal MDS stem cell. The hypomethylating agents azacitidine and decitabine are frequently used in bone marrow failure clinics. Azacitidine improves blood counts and survival in MDS, compared to best supportive care. Azacitidine is usually administered subcutaneously, daily for 7 days, at 4-week intervals, for at least four cycles before assessing for response. Overall, generally improved blood counts with a decrease in transfusion requirements occurred in ~50% of patients in published trials. Response is dependent on continued drug administration, and most patients eventually become refractory to drug intervention and experience recurrent cytopenias or progression to AML. Decitabine is closely related to azacitidine; 30–50% of patients show responses in blood counts, with a duration of response of almost a year. Decitabine is usually administered by continuous intravenous infusion in regimens of varying doses and durations of 3–10 days in repeating cycles. The major toxicity of azacitidine and decitabine is myelosuppression, leading to worsening blood counts. Hypomethylating agents are frequently used in the high-risk patient who is not a candidate for stem cell transplant. In the lower risk patient, they are also effective, but alternative therapies should be considered.

Lenalidomide, a thalidomide derivative with a more favorable toxicity profile, is particularly effective in reversing anemia in MDS patients with 5q- syndrome; not only do a high proportion of these

TABLE 102-6 Revised International Prognostic Scoring System (IPSS-R)

1. New marrow blast categories
≤2%, >2%–<5%, 5–10%, >10–30%
2. Refined cytogenetic abnormalities and risk groups
16 (vs 6) specific abnormalities, 5 (vs 3) subgroups^a
3. Evaluation of depth of cytopenias^b
Clinically and statistically relevant cut points used
4. Inclusion of differentiating features
Age, performance status, serum ferritin, LDH; β₂-microglobulin
5. Prognostic model with 5 (vs 4) risk categories
Improved predictive power

^aGood, normal, –Y, del(5q), del(20q); poor, complex (≥3 abnormalities) or chromosome 7 abnormalities; intermediate, all other abnormalities. ^bCytopenias at baseline, cut points: hemoglobin <80, 80–<100, or ≥100 g/L; platelet count <50, 50–100, or ≥100,000/µL, and absolute neutrophil count <800 versus ≥800/µL.

Abbreviation: LDH, lactate dehydrogenase.

patients become transfusion independent with normal or near-normal hemoglobin levels, but their cytogenetics also become normal. The drug has many biologic activities, and it is unclear which is critical for clinical efficacy. Lenalidomide is administered orally. Most patients will improve within 3 months of initiating therapy. Toxicities include myelosuppression (worsening thrombocytopenia and neutropenia, necessitating blood count monitoring) and an increased risk of deep vein thrombosis and pulmonary embolism.

Immunosuppression also may produce sustained independence from transfusion and improve survival. ATG, cyclosporine, and the anti-CD52 monoclonal antibody alemtuzumab are especially effective in younger MDS patients (<60 years old) with more favorable IPSS. In a consortium retrospective review, about 50% of patients with mainly refractory anemia responded to ATG, usually combined with cyclosporine, particularly patients with hypocellular marrow.

HGFs can improve blood counts but, as in most other marrow failure states, have been most beneficial to patients with the least severe pancytopenia. EPO alone or in combination with G-CSF can improve hemoglobin levels, particularly in those with low serum EPO levels who have no or a modest need for transfusions. Survival may be enhanced by EPO and amelioration of anemia. G-CSF treatment alone failed to improve survival in a controlled trial. Thrombopoietin mimetics appear to improve platelet counts in some MDS patients, with no clear evidence that they increase the rate of leukemic transformation.

New drugs for MDS are entering the clinic or are in late development. Luspatercept, which affects transforming growth factor β -mediated suppression of erythropoiesis, has been approved by the FDA for anemia in MDS. Novel targeted therapies in trials include inhibitors of hypoxia-inducible factor and spliceosome genes, drugs that act to restore TP53 activity, and venetoclax, an inhibitor of the bcl2 protein that increases programmed cell death (and is approved for use or employed off-label in other hematologic malignancies).

The same principles of supportive care described for aplastic anemia apply to MDS. Many patients will be anemic for years. RBC transfusion support should be accompanied by iron chelation to prevent secondary hemochromatosis.

MYELOPHTHISIS ANEMIAS

Fibrosis of the bone marrow (see Fig. 100-2), usually accompanied by a characteristic blood smear picture called *leukoerythroblastosis*, can occur as a primary hematologic disease, called *myelofibrosis* or *myeloid metaplasia* (Chap. 103), and as a secondary process, called *myelophthisis*. Myelophthisis, or secondary myelofibrosis, is reactive. Fibrosis can be a response to invading tumor cells, usually an epithelial cancer of breast, lung, or prostate origin or neuroblastoma. Marrow fibrosis may occur with infection of mycobacteria (both *Mycobacterium tuberculosis* and *Mycobacterium avium*), fungi, or HIV and in sarcoidosis. Intracellular lipid deposition in Gaucher disease and obliteration of the marrow space related to absence of osteoclast remodeling in congenital osteopetrosis also can produce fibrosis. Secondary myelofibrosis is a late consequence of radiation therapy or treatment with radiomimetic drugs. Usually the infectious or malignant underlying processes are obvious. Marrow fibrosis can also be a feature of a variety of hematologic syndromes, especially chronic myeloid leukemia, multiple myeloma, lymphomas, myeloma, and hairy cell leukemia.

The pathophysiology has three distinct features: proliferation of fibroblasts in the marrow space (myelofibrosis); the extension of hematopoiesis into the long bones and into extramedullary sites, usually the spleen, liver, and lymph nodes (myeloid metaplasia); and ineffective erythropoiesis. The etiology of the fibrosis is unknown but most likely involves dysregulated production of growth factors: platelet-derived growth factor and transforming growth factor β have been implicated. Abnormal regulation of other hematopoietins would lead to localization of blood-producing cells in nonhematopoietic tissues and uncoupling of the usually balanced processes of stem cell proliferation and differentiation. Myelofibrosis is remarkable for

pancytopenia despite very large numbers of circulating hematopoietic progenitor cells.

Anemia is dominant in secondary myelofibrosis, usually normocytic and normochromic. The diagnosis is suggested by the characteristic leukoerythroblastotic smear (see Fig. 100-1). Erythrocyte morphology is highly abnormal, with circulating nucleated RBCs, teardrops, and shape distortions. WBC numbers are often elevated, sometimes mimicking a leukemoid reaction, with circulating myelocytes, promyelocytes, and myeloblasts. Platelets may be abundant and are often of giant size. Inability to aspirate the bone marrow, the characteristic “dry tap,” can allow a presumptive diagnosis in the appropriate setting before the biopsy is decalcified.

The course of secondary myelofibrosis is determined by its etiology, usually a metastatic tumor or an advanced hematologic malignancy. Treatable causes must be excluded, especially tuberculosis and fungus. Transfusion support can relieve symptoms.

FURTHER READING

- ARBER DA et al: The 2016 revision to the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia. *Blood* 127:2391, 2016.
- ATTALAH E et al: Comparison of patient age groups in transplantation for myelodysplastic syndrome: The Medicare coverage with evidence development study. *JAMA Oncol* 6:486, 2020.
- DEZERN AE et al: Haplodidential BMT for severe aplastic anemia with intensive GVHD prophylaxis including posttransplant cyclophosphamide. *Blood Adv* 4: 1770, 2020.
- GREENBERG PL et al: Revised international prognostic scoring system for myelodysplastic syndromes. *Blood* 120:2454, 2012.
- MUSTJOKI S, YOUNG NS: Somatic mutations in “benign” disease. *N Engl J Med* 384:2039, 2021.
- OGAWA S: Genetics of MDS. *Blood* 133:1049, 2019.
- PLATZBECKER U: Treatment of MDS. *Blood* 133:1096, 2019.
- SALLMAN DA, LIST AF: The central role of inflammatory signaling in the pathogenesis of myelodysplastic syndromes. *Blood* 133:1039, 2019.
- TOWNSLEY DM et al: Eltrombopag added to standard immunosuppression for aplastic anemia. *N Engl J Med* 376:1540, 2017.
- YOSHIZATO T et al: Somatic mutations and clonal hematopoiesis in aplastic anemia. *N Engl J Med* 373:35, 2015.
- YOUNG NS: Aplastic anemia. *N Engl J Med* 379:1643, 2018.

103

Polycythemia Vera and Other Myeloproliferative Neoplasms

Jerry L. Spivak



The World Health Organization (WHO) classification of the chronic myeloproliferative neoplasms (MPNs) includes eight disorders, some of which are rare or poorly characterized (Table 103-1) but all of which share an origin in a hematopoietic cell; overproduction of one or more of the formed elements of the blood without significant dysplasia; and a predilection to extramedullary hematopoiesis, myelofibrosis, and transformation at varying rates to acute leukemia. Within this broad classification, however, significant phenotypic heterogeneity exists. Some diseases such as chronic myelogenous leukemia (CML), chronic neutrophilic leukemia (CNL), and chronic eosinophilic leukemia (CEL) express primarily a myeloid phenotype, whereas in other diseases, such as polycythemia vera (PV), primary myelofibrosis (PMF), and essential thrombocythemia (ET), erythroid or megakaryocytic

TABLE 103-1 World Health Organization Classification of Chronic Myeloproliferative Neoplasms

Chronic myeloid leukemia, <i>BCR-ABL</i> -positive
Chronic neutrophilic leukemia
Chronic eosinophilic leukemia, not otherwise specified
Polycythemia vera
Primary myelofibrosis
Essential thrombocythemia
Mastocytosis
Myeloproliferative neoplasms, unclassifiable

hyperplasia predominates. The latter three disorders, in contrast to the former three, also appear capable of transforming into each other.

Such phenotypic heterogeneity has a genetic basis; CML is the consequence of the balanced translocation between chromosomes 9 and 22 (t[9;22][q34;11]); CNL has been associated with a t(15;19) translocation; and CEL occurs with a deletion or balanced translocations involving the *PDGFRα* gene. By contrast, PV, PMF, and ET are characterized by driver mutations that directly or indirectly constitutively activate JAK2, a tyrosine kinase essential for the function of the erythropoietin and thrombopoietin receptors and also utilized by the granulocyte colony-stimulating factor receptor. This important distinction is reflected in the natural histories of CML, CNL, and CEL, which are usually measured in years, with a high rate of leukemic transformation. The natural histories of PV, PMF, and ET, by contrast, are usually measured in decades, and transformation to acute leukemia is uncommon in the absence of chemotherapy. This chapter focuses only on PV, PMF, and ET because their clinical features and driver mutation overlap are substantial, although their disease duration varies.

The other chronic MPNs will be discussed in *Chaps. 105* and *110*.

POLYCYTHEMIA VERA

PV is a clonal hematopoietic stem cell disorder in which phenotypically normal red cells, granulocytes, and platelets accumulate in the absence of a recognizable physiologic stimulus. The most common of the MPNs, PV occurs in 2.5 per 100,000 persons, sparing no adult age group and increasing with age to rates >10/100,000. Familial transmission is infrequent, and women under age 50 predominate among sporadic cases.

ETIOLOGY

Nonrandom chromosome abnormalities such as deletion 20q and deletion 13q or trisomy 9 occur in up to 30% of untreated PV patients, but unlike CML, no consistent cytogenetic abnormality has been associated with the disorder. However, a mutation in the autoinhibitory pseudokinase domain of the tyrosine kinase JAK2 that replaces valine with phenylalanine (V617F), causing constitutive kinase activation, has a central role in PV pathogenesis.

JAK2 is a member of an evolutionarily well-conserved, nonreceptor tyrosine kinase family and serves as the cognate tyrosine kinase for the erythropoietin and thrombopoietin receptors. It also functions as an obligate chaperone for these receptors in the Golgi apparatus and is responsible for their cell-surface expression. The conformational change induced in the erythropoietin and thrombopoietin receptors following binding to their respective cognate ligands, erythropoietin or thrombopoietin, leads to JAK2 autophosphorylation, receptor phosphorylation, and phosphorylation of proteins involved in cell proliferation, differentiation, and resistance to apoptosis. Transgenic animals lacking JAK2 die as embryos from severe anemia. Constitutive activation of JAK2, on the other hand, explains the erythropoietin hypersensitivity, erythropoietin-independent erythroid colony formation, rapid terminal differentiation, increased Bcl-X_L expression, and apoptosis resistance in the absence of erythropoietin that characterize the *in vitro* behavior of PV erythroid progenitor cells.

More than 95% of PV patients express this mutation, as do ~50% of PMF and ET patients. Importantly, the JAK2 gene is located on

the short arm of chromosome 9, and loss of heterozygosity on chromosome 9p involving the segment containing the JAK2 locus over time due to mitotic recombination (uniparental disomy) is the most common cytogenetic abnormality in PV. Loss of heterozygosity in this region leads to homozygosity for JAK2 V617F and occurs in ~60% of PV patients and to a lesser extent in PMF but is rare in ET. Most PV patients who do not express JAK2 V617F express a mutation in exon 12 of the gene and are not clinically different from those who do, with the exception of a higher frequency of isolated erythrocytosis, nor do JAK2 V617F heterozygotes differ clinically from homozygotes. Importantly, the predisposition to acquire JAK2 mutations appears to be associated with a specific JAK2 gene haplotype, GGCC. JAK2 V617F is the basis for many of the phenotypic and biochemical characteristics of PV such as increased blood cell production and increased inflammatory cytokine production; however, it cannot solely account for the entire PV phenotype and is probably not the initiating lesion in any of the MPNs. First, PV patients with the same phenotype and documented clonal disease can have mutations in *LNK*, a JAK2 inhibitor, or rarely, calreticulin (*CALR*), an ER chaperone. Second, ET and PMF patients have the same mutation but different clinical phenotypes. Third, familial PV can occur without the mutation, even when other members of the same family express it. Fourth, inhibition of JAK2 V617F-expressing hematopoietic progenitor cells by the nonspecific JAK1/2 kinase inhibitor, ruxolitinib, does not affect the behavior of the involved hematopoietic stem cells. Finally, in some JAK2 V617F-positive PV or ET patients, acute leukemia can occur in a JAK2 V617F-negative progenitor cell, suggesting the presence of an ancestral precursor cell.

CLINICAL FEATURES

Although PV is a panmyelopathy, isolated thrombocytosis, leukocytosis, or splenomegaly may be its initial presenting manifestation, but most often, the disorder is first recognized by the incidental discovery of a high hemoglobin, hematocrit, or red cell count. With the exception of aquagenic pruritus, or erythromelalgia, no symptoms distinguish PV from other causes of erythrocytosis.

Uncontrolled erythrocytosis causes hyperviscosity, leading to neurologic symptoms such as vertigo, tinnitus, headache, visual disturbances, and transient ischemic attacks (TIAs). Systolic hypertension is also a feature of the red cell mass elevation. In some patients, venous or arterial thrombosis may be the presenting manifestation of PV. Any vessel can be affected, but cerebral, cardiac, and mesenteric vessels are most commonly involved. Hepatic venous thrombosis (Budd-Chiari syndrome) is particularly common in young women and may be catastrophic if sudden and complete obstruction of the hepatic vein occurs. Indeed, PV should be suspected in any patient who develops hepatic vein thrombosis, since this is the only type of thrombosis associated with JAK2 V617F expression. Digital ischemia, easy bruising, epistaxis, acid-peptic disease, or gastrointestinal hemorrhage may occur due to vascular stasis or thrombocytosis. In the latter instance, absorption and proteolysis of high-molecular-weight von Willebrand multimers by the large platelet mass cause acquired von Willebrand's disease. Erythema, burning, and pain in the extremities, a symptom complex known as erythromelalgia, is another complication of thrombocytosis in PV due to increased platelet stickiness. Given the large turnover of hematopoietic cells, hyperuricemia with secondary gout, uric acid stones, and symptoms due to hypermetabolism can also complicate the disorder.

DIAGNOSIS

When PV presents with erythrocytosis in combination with leukocytosis, thrombocytosis, or splenomegaly or any combination of these, the diagnosis is apparent. However, when patients present with an elevated hemoglobin, hematocrit, or red cell count alone, the diagnostic evaluation is more complex because of the many diagnostic possibilities (*Table 103-2*). Furthermore, unless the hemoglobin level is ≥20 g/dL (hematocrit ≥60%), it is not possible to distinguish true erythrocytosis from disorders causing plasma volume contraction. This is because uniquely in PV, in contrast to other causes of true erythrocytosis, there is expansion of the plasma volume, which can mask the elevated red cell mass, particularly in women; thus, red cell mass and plasma

TABLE 103-2 Causes of Erythrocytosis**Relative Erythrocytosis**

Hemoconcentration secondary to dehydration, diuretics, ethanol abuse, androgens, or tobacco abuse

Absolute Erythrocytosis

Hypoxia	Tumors
Carbon monoxide intoxication	Hypernephroma
High-oxygen-affinity hemoglobins	Hepatoma
High altitude	Cerebellar hemangioblastoma
Pulmonary disease	Uterine myoma
Right-to-left cardiac or vascular shunts	Adrenal tumors
Sleep apnea syndrome	Meningioma
Hepatopulmonary syndrome	Pheochromocytoma
Renal Disease	Drugs
Renal artery stenosis	Androgens
Focal sclerosing or membranous glomerulonephritis	Recombinant erythropoietin
Postrenal transplantation	Familial (with normal hemoglobin function)
Renal cysts	Erythropoietin receptor mutations
Bartter's syndrome	VHL mutations (Chuvash polycythemia)
	2,3-BPG mutation
	PHD2 and HIF2α mutations
	Polycythemia vera

Abbreviations: 2,3-BPG, 2,3-bisphosphoglycerate; VHL, von Hippel-Lindau.

volume determinations are necessary to establish the presence of an absolute erythrocytosis and distinguish this from relative erythrocytosis due to a reduction in plasma volume alone (also known as *stress* or *spurious erythrocytosis* or *Gaisböck's syndrome*). **Figure 63-18** illustrates a diagnostic algorithm for the evaluation of suspected erythrocytosis. Assay for JAK2 mutations in the presence of a normal arterial oxygen saturation provides an alternative diagnostic approach to erythrocytosis when red cell mass and plasma volume determinations are not available; a normal serum erythropoietin level does not exclude the presence of PV, but an elevated erythropoietin level is more consistent with a secondary cause for the erythrocytosis.

Other laboratory studies that may aid in diagnosis include the red cell count, mean corpuscular volume, and red cell distribution width (RDW), particularly when the hematocrit or hemoglobin levels are less than 60% or 20 g/dL, respectively. Only three situations cause microcytic erythrocytosis: β-thalassemia trait, hypoxic erythrocytosis, and PV. With β-thalassemia trait, the RDW is usually normal, whereas with hypoxic erythrocytosis and PV, the RDW may be elevated due to associated iron deficiency. Today, however, the assay for JAK2 V617F has superseded other tests for establishing the diagnosis of PV. Of course, in patients with associated acid-peptic disease, occult gastrointestinal bleeding may lead to a presentation with hypochromic, microcytic anemia, masking the presence of PV.

A bone marrow aspirate and biopsy provide no specific diagnostic information because these may be normal or indistinguishable from ET or PMF. Similarly, no specific cytogenetic abnormality is associated with the disease, and the absence of a cytogenetic marker does not exclude the diagnosis.

COMPLICATIONS

Many of the clinical complications of PV relate directly to the increase in blood viscosity associated with red cell mass elevation and indirectly to the increased turnover of red cells, leukocytes, and platelets with the attendant increase in uric acid and inflammatory cytokine production. The latter appears to be responsible for constitutional symptoms. Peptic ulcer disease can also be due to *Helicobacter pylori* infection, the incidence of which is increased in PV, while the pruritus associated with this disorder may be a consequence of mast cell activation by JAK2 V617F. A sudden increase in spleen size can be associated with

painful splenic infarction. Myelofibrosis appears to be part of the natural history of the disease but is a reactive, reversible process that does not itself impede hematopoiesis and by itself has no prognostic significance. In ~15% of patients, however, myelofibrosis is associated with hematopoietic stem cell failure, manifested by substantial extramedullary hematopoiesis in the liver and spleen and transfusion-dependent anemia. The organomegaly can cause significant mechanical discomfort, portal hypertension, and progressive cachexia. Although the incidence of acute myeloid leukemia is increased in PV, the incidence of acute leukemia in patients not exposed to chemotherapy or radiation therapy is low. Interestingly, chemotherapy, including hydroxyurea, has been associated with acute leukemia in JAK2 V617F-negative stem cells in some PV patients. *Erythromelalgia* is a curious syndrome of unknown etiology associated with thrombocytosis, primarily involving the lower extremities and usually manifested by erythema, warmth, and pain of the affected appendage and occasionally digital infarction. It occurs with a variable frequency and is usually responsive to salicylates. Some of the central nervous system symptoms observed in patients with PV, such as ocular migraine, appear to represent a variant of erythromelalgia.

Left uncontrolled, erythrocytosis can lead to thrombosis involving vital organs such as the liver, heart, brain, or lungs. Patients with massive splenomegaly are particularly prone to thrombotic events because the associated increase in plasma volume masks the true extent of the red cell mass elevation measured by the hematocrit or hemoglobin level. A "normal" hematocrit or hemoglobin level in a PV patient with massive splenomegaly should be considered indicative of an elevated red cell mass until proven otherwise.

TREATMENT

Polycythemia Vera

PV is generally an indolent disorder, the clinical course of which is measured in decades, and its management should reflect its tempo. Thrombosis due to erythrocytosis is the most significant complication and often the presenting manifestation; maintenance of the hemoglobin level at ≤140 g/L (14 g/dL; hematocrit <45%) in men and ≤120 g/L (12 g/dL; hematocrit <42%) in women is mandatory to avoid thrombotic complications. Phlebotomy serves initially to reduce hyperviscosity by reducing the red cell mass to normal while further expanding the plasma volume. Periodic phlebotomies thereafter serve to maintain the red cell mass within the normal range and induce a state of iron deficiency that prevents accelerated reexpansion of the red cell mass. In most PV patients, once an iron-deficient state is achieved, phlebotomy is usually only required at 3-month intervals. Neither phlebotomy nor iron deficiency increases the platelet count relative to the effect of the disease itself, and neither thrombocytosis nor leukocytosis is correlated with thrombosis in PV, in contrast to the strong correlation between erythrocytosis and thrombosis. The use of salicylates to prevent thrombosis in PV patients is not only potentially harmful if the red cell mass is not controlled by phlebotomy but also an unproven remedy, particularly in patients over age 70.

Anticoagulation is indicated when a thrombosis has occurred, and the newer oral anticoagulants may be preferable to a vitamin K antagonist since they do not require monitoring. Asymptomatic hyperuricemia (<10 mg/dL) requires no therapy, but allopurinol should be administered to avoid further elevation of the uric acid when chemotherapy is used to reduce splenomegaly or leukocytosis or to treat pruritus. Generalized pruritus intractable to antihistamines or antidepressants such as doxepin can be a major problem in PV; the JAK1/2 inhibitor ruxolitinib, pegylated interferon α (IFN-α), psoralens with ultraviolet light in the A range (PUVA) therapy, and hydroxyurea are other methods of palliation. Asymptomatic thrombocytosis requires no therapy unless the platelet count is sufficiently high to cause bleeding due to acquired von Willebrand's disease, but bleeding in this situation is not usually spontaneous and is responsive to tranexamic acid or ε-aminocaproic acid. Symptomatic

splenomegaly can be treated with either ruxolitinib or pegylated IFN- α . Pegylated IFN- α has the advantage over recombinant IFN- α of being better tolerated and requiring only weekly administration and produced complete hematologic and molecular remissions in ~20% of PV patients; its role in this disorder is currently under investigation. Anagrelide, a phosphodiesterase inhibitor, can reduce the platelet count and, if tolerated, is preferable to hydroxyurea because it lacks marrow toxicity and is also protective against venous thrombosis while hydroxyurea is not.

A reduction in platelet number may be necessary for the treatment of erythromelalgia or ocular migraine if salicylates are not effective or if the platelet count is sufficiently high to increase the risk of hemorrhage but only to the degree that symptoms are alleviated. Alkylating agents and radioactive sodium phosphate (^{32}P) are leukemogenic in PV, and their use should be avoided. If a cytotoxic agent must be used, hydroxyurea is preferred, but this drug does not prevent either thrombosis or myelofibrosis in PV, is itself leukemogenic, and should be used for as short a time as possible. Previously, PV patients with massive splenomegaly unresponsive to reduction by chemotherapy or interferon required splenectomy. However, with the introduction of the nonspecific JAK2 inhibitor ruxolitinib, it has been possible in the majority of patients with PV complicated by myelofibrosis and myeloid metaplasia to reduce spleen size while at the same time alleviating constitutional symptoms and pruritus due to cytokine release and reducing the phlebotomy requirement. However, in contrast to PMF, these patients have a more chronic course. In contrast to other malignancies, PV patients have a low rate of mutation accumulation, and the acquisition of deleterious mutations such as *TP53* mutations as detected by next-generation sequencing is usually associated with leukemic transformation. Since hydroxyurea antagonizes *TP53* and also causes del17p, leading to *TP53* haploinsufficiency, its use should be constrained in PV.

Ruxolitinib has also been demonstrated in a phase 3 clinical trial to be effective in PV patients without myelofibrosis who are intolerant or refractory to hydroxyurea or best available supportive therapy. In some patients with end-stage disease, pulmonary hypertension may develop due to fibrosis or extramedullary hematopoiesis. A role for bone marrow transplantation, either allogeneic or haploidentical, in PV has not been defined.

Most patients with PV can live long lives without functional impairment when their red cell mass is effectively managed with phlebotomy alone. Chemotherapy is never indicated to control the red cell mass in PV, but when venous access is an issue, ruxolitinib or pegylated interferon is the preferred therapy.

■ PRIMARY MYELOFIBROSIS

Chronic PMF (other designations include *idiopathic myelofibrosis*, *agnogenic myeloid metaplasia*, or *myelofibrosis with myeloid metaplasia*) is a clonal hematopoietic stem cell disorder associated with mutations in *JAK2*, *MPL*, or *CALR* and characterized by marrow fibrosis, extramedullary hematopoiesis, and splenomegaly. PMF is the least common MPN, and establishing its diagnosis in the absence of a specific clonal marker is difficult because myelofibrosis and splenomegaly are also features of both PV and CML. Furthermore, myelofibrosis and splenomegaly also occur in a variety of benign and malignant disorders (Table 103-3), many of which are amenable to specific therapies not effective in PMF. In contrast to the other MPNs and so-called acute or malignant myelofibrosis, which can occur at any age, PMF primarily afflicts men in their sixth decade or later.

■ ETIOLOGY

Nonrandom chromosome abnormalities such as 9p-, 20q-, 13q-, trisomy 8 or 9, or partial trisomy 1q are common in PMF, but no cytogenetic abnormality specific to the disease has been identified. *JAK2* V617F is present in ~55% of PMF patients, and mutations in the thrombopoietin receptor, *MPL*, occur in ~4%. Most of the rest have mutations in the calreticulin gene (*CALR*) that alter the carboxy-terminal portion of the protein, permitting it to bind and activate *MPL*.

TABLE 103-3 Disorders Causing Myelofibrosis

MALIGNANT	NONMALIGNANT
Acute leukemia (lymphocytic, myelogenous, megakaryocytic)	HIV infection
Chronic myeloid leukemia	Hyperparathyroidism
Hairy cell leukemia	Renal osteodystrophy
Hodgkin's disease	Systemic lupus erythematosus
Primary myelofibrosis	Tuberculosis
Lymphoma	Vitamin D deficiency
Multiple myeloma	Thorium dioxide exposure
Myelodysplasia	Gray platelet syndrome
Metastatic carcinoma	
Polycythemia vera	
Systemic mastocytosis	

The degree of myelofibrosis and the extent of extramedullary hematopoiesis are not related. Fibrosis in this disorder is associated with overproduction of transforming growth factor β and tissue inhibitors of metalloproteinases, while osteosclerosis is associated with overproduction of osteoprotegerin, an osteoclast inhibitor. Marrow angiogenesis occurs due to increased production of vascular endothelial growth factor. Importantly, fibroblasts in PMF are polyclonal and not part of the neoplastic clone but can be induced by it to produce inflammatory cytokines.

■ CLINICAL FEATURES

No signs or symptoms are specific for PMF. Many patients are asymptomatic at presentation, and the disease is often detected by the discovery of splenic enlargement and/or abnormal blood counts during a routine examination. In contrast to its companion MPN, night sweats, fatigue, and weight loss are common presenting complaints. A blood smear will show the characteristic features of extramedullary hematopoiesis: teardrop-shaped red cells, nucleated red cells, myelocytes, and promyelocytes; myeloblasts may also be present (Fig. 103-1). Anemia, usually mild initially, is common, whereas the leukocyte and platelet counts are either normal or increased, but either can be depressed. Mild hepatomegaly may accompany the splenomegaly but is unusual in its absence; isolated lymphadenopathy should suggest another diagnosis. Both serum lactate dehydrogenase and alkaline phosphatase levels can be elevated. Marrow is usually inaspirable due to the myelofibrosis (Fig. 103-2), and bone x-rays may reveal osteosclerosis. Exuberant extramedullary hematopoiesis can cause ascites; portal, pulmonary, or

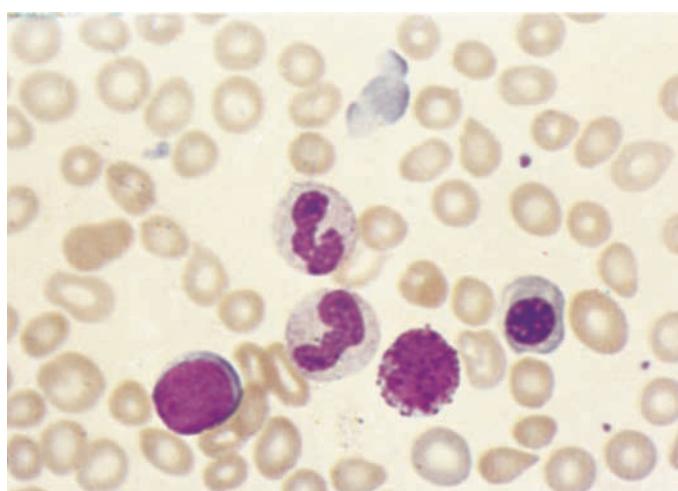


FIGURE 103-1 Teardrop-shaped red blood cells indicative of membrane damage from passage through the spleen, a nucleated red blood cell, and immature myeloid cells indicative of extramedullary hematopoiesis are noted. This peripheral blood smear is related to any cause of extramedullary hematopoiesis.

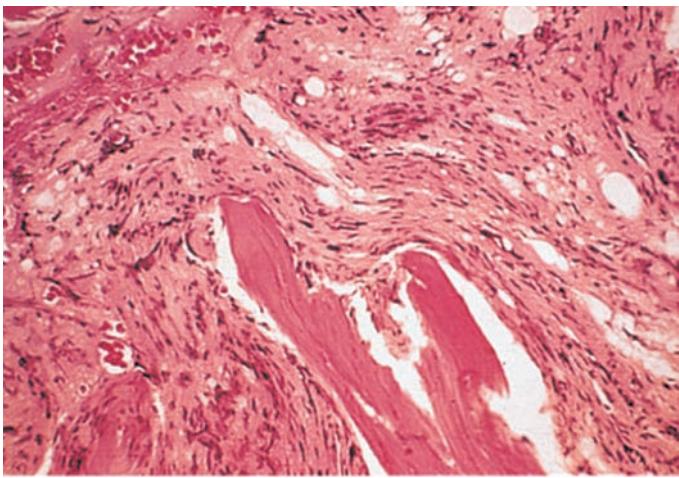


FIGURE 103-2 This marrow section shows the marrow cavity replaced by fibrous tissue composed of reticulin fibers and collagen. When this fibrosis is due to a primary hematologic process, it is called *myelofibrosis*. When the fibrosis is secondary to a tumor or a granulomatous process, it is called *myelophthisis*.

intracranial hypertension; intestinal or ureteral obstruction; pericardial tamponade; spinal cord compression; or skin nodules. Splenic enlargement can be sufficiently rapid to cause splenic infarction with fever and pleuritic chest pain. Hyperuricemia and secondary gout may ensue.

■ DIAGNOSIS

While the clinical picture described above is characteristic of PMF, all of these clinical features can be observed in PV or CML. Massive splenomegaly commonly masks erythrocytosis in PV, and reports of intraabdominal thrombosis in PMF most likely represent instances of unrecognized PV. In some PMF patients, erythrocytosis has developed during the course of the disease. Furthermore, because many other disorders have features that overlap with PMF but respond to distinctly different therapies, the diagnosis of PMF is one of exclusion, which requires that the disorders listed in Table 103-3 be ruled out.

The presence of teardrop-shaped red cells, nucleated red cells, myelocytes, and promyelocytes establishes the presence of extramedullary hematopoiesis, while the presence of leukocytosis, thrombocytosis with large and bizarre platelets, and circulating myelocytes suggests the presence of an MPN as opposed to a secondary form of myelofibrosis (Table 103-3). Marrow is usually inaspirable due to increased marrow reticulin, but marrow biopsy will reveal a hypercellular marrow with trilineage hyperplasia and, in particular, increased numbers of megakaryocytes in clusters and with large, dysplastic nuclei. However, there are no characteristic bone marrow morphologic abnormalities that distinguish PMF from the other MPNs. Splenomegaly due to extramedullary hematopoiesis may be sufficiently massive to cause portal hypertension and variceal formation. In some patients, exuberant extramedullary hematopoiesis can dominate the clinical picture. An intriguing feature of PMF is the occurrence of autoimmune abnormalities such as immune complexes, antinuclear antibodies, rheumatoid factor, or a positive Coombs' test. Whether these represent a host reaction to the disorder or are involved in its pathogenesis is unknown. Cytogenetic analysis of the blood is useful both to exclude CML and for prognostic purposes because the development of complex karyotype abnormalities portends a poor prognosis in PMF. For unknown reasons, the number of circulating CD34+ cells is markedly increased in PMF ($>15,000/\mu\text{L}$) compared to the other MPNs, unless they too develop extramedullary hematopoiesis.

Importantly, ~55% of PMF patients, like patients with its companion MPNs, express the *JAK2* V617F mutation, often as homozygotes. Such patients are usually older and have higher hematocrits than patients with *MPL* (4%) or *CALR* (36%) mutations; PMF patients expressing an *MPL* mutation tend to be more anemic and have lower leukocyte counts than *JAK2* V617F-positive patients. Somatic mutations (due to deletions [type 1] or insertions [type 2]) in exon 9 of *CALR* have been

TABLE 103-4 Three Current Scoring Systems for Estimating Prognosis in PMF Patients

RISK FACTOR	IPSS (2009) ^a	DIPSS (2010) ^b	DIPSS PLUS (2011) ^c
Anemia ($<10 \text{ g/dL}$)	X	X	X
Leukocytosis ($>25,000/\mu\text{L}$)	X	X	X
Peripheral blood blasts ($\geq 1\%$)	X	X	X
Constitutional symptoms	X	X	X
Age ($>65 \text{ years}$)	X	X	X
Unfavorable karyotype			X
Platelet count ($<100,000/\mu\text{L}$)			X
Transfusion dependence			X

^aBlood 113:2895, 2009. ^bBlood 115:1703, 2010. ^cJ Clin Oncol 29:392, 2011.

Note: The Dynamic International Prognostic Scoring System (DIPSS) was developed to determine if the International Prognostic Scoring System (IPSS) risk factors identified as important for survival at the time of primary myelofibrosis (PMF) diagnosis could also be used for risk stratification following their acquisition during the course of the disease. One point is assigned to each risk factor for IPSS scoring. For DIPSS, the same is true, but anemia is assigned 2 points. The DIPSS Plus scoring system represents recognition that the addition of unfavorable karyotype, thrombocytopenia, and transfusion dependence improved the DIPSS risk stratification system for which additional points are assigned (Table 103-5). More recent studies suggest that mutational analysis of the *ASXL1*, *EZH2*, *SRSF2*, and *IDH1/2* genes further improves risk stratification for survival and leukemic transformation (Leukemia 27:1861, 2013), as can cytogenetic abnormalities (Leukemia 32:1631, 2018). These prognostic scoring systems are not accurate for risk assessment in polycythemia vera or essential thrombocythosis patients who have developed myelofibrosis (Haematologica 99:e55, 2014).

found in a majority of patients with PMF who lack mutations in either *JAK2* or *MPL*. In some studies, type 1 mutations, the most common *CALR* mutation in PMF, had a survival advantage compared to *JAK2* or *MPL* mutations but not with respect to leukemic transformation. PMF patients who lack a known MPN driver mutation appear to have the worst prognosis.

■ COMPLICATIONS

Survival in PMF varies according to specific risk factors at diagnosis (Tables 103-4 and 103-5) but is shorter than in PV and ET patients. The natural history of PMF is one of increasing marrow failure with transfusion-dependent anemia and increasing organomegaly due to extramedullary hematopoiesis. As with CML, PMF can evolve from a chronic to an accelerated phase with constitutional symptoms and increasing marrow failure. About 10% of patients spontaneously transform to an aggressive form of acute leukemia for which therapy is usually ineffective. Additional important prognostic factors for disease acceleration during the course of PMF include the presence of complex cytogenetic abnormalities, thrombocytopenia, and transfusion-dependent anemia. Mutations in the *ASXL1*, *EZH2*, *SRSF2*, and *IDH1/2* genes have been identified as risk factors for early death or transformation to acute leukemia, as have complex cytogenetic abnormalities, and have proved to be more useful for PMF risk assessment than clinical scoring systems.

TABLE 103-5 IPSS and DIPSS Risk Stratification Systems

RISK CATEGORIES ^a	NUMBER OF RISK FACTORS		
	IPSS	DIPSS	DIPSS PLUS
Low	0	0	0
Intermediate-1	1	1–2	1
Intermediate-2	2	3–4	2–3
High	≥ 3	>4	4–6

^aThe corresponding survival curves for each risk category can be found in the references cited in the footnotes of Table 103-4.

Abbreviations: DIPSS, Dynamic International Prognostic Scoring System; IPSS, International Prognostic Scoring System.

TREATMENT

Primary Myelofibrosis

No specific therapy exists for PMF. The causes for anemia are multifarious and include ineffective erythropoiesis uncompensated by splenic extramedullary hematopoiesis, hemodilution due to splenomegaly, splenic sequestration, blood loss secondary to thrombocytopenia or portal hypertension, folic acid deficiency, systemic inflammation, and autoimmune hemolysis. Neither recombinant erythropoietin nor androgens such as danazol have proven to be consistently effective as therapy for anemia. Erythropoietin may worsen splenomegaly and will be ineffective if the serum erythropoietin level is >125 mU/L. Given the inflammatory milieu that characterizes PMF, glucocorticoids can ameliorate anemia as well as constitutional symptoms such as fever, chills, night sweats, anorexia, and weight loss, and combining these with low-dose thalidomide has proved effective as well. Thrombocytopenia can be due to impaired marrow function, splenic sequestration, or autoimmune destruction and may also respond to low-dose thalidomide and prednisone.

Splenomegaly is by far the most distressing and intractable problem for PMF patients, causing abdominal pain, portal hypertension, easy satiety, and cachexia, whereas surgical removal of a massive spleen is associated with significant postoperative complications including mesenteric venous thrombosis, hemorrhage, rebound leukocytosis and thrombocytosis, and hepatic extramedullary hematopoiesis with no amelioration of either anemia or thrombocytopenia when present. For unexplained reasons, splenectomy also increases the risk of blastic transformation.

Splenic irradiation is, at best, temporarily palliative and associated with a significant risk of neutropenia, infection, and subsequent operative hemorrhage if splenectomy is attempted. Allopurinol can control significant hyperuricemia, and bone pain can be alleviated by local irradiation. Pegylated IFN- α can ameliorate fibrosis in early PMF, but in advanced disease, it may exacerbate the bone marrow failure. The JAK2 inhibitor ruxolitinib has proved effective in reducing splenomegaly and alleviating constitutional symptoms in a majority of advanced PMF patients while possibly prolonging survival, although it usually does not significantly influence the JAK2 V617F neutrophil allele burden. Although anemia and thrombocytopenia are its major side effects, these are dose-dependent, and with time, anemia stabilizes and thrombocytopenia may improve. Fedratinib, a new tyrosine kinase inhibitor with anti-FLT3 activity, has proved useful in patients with disease refractory to ruxolitinib.

In some patients, hypomethylating agents such as azacytidine or decitabine in combination with high-dose ruxolitinib have been used to control the disease or prepare patients for bone marrow transplantation. Transformation to acute leukemia in PMF, like PV or ET, is usually refractory to treatment.

Allogeneic bone marrow transplantation is the only curative treatment for PMF and should be considered in younger patients and older patients with high-risk disease; nonmyeloablative conditioning regimens permit hematopoietic cell transplantation to be extended to older individuals.

ESSENTIAL THROMBOCYTOSIS

ET (other designations include *essential thrombocythemia*, *idiopathic thrombocytosis*, *primary thrombocytosis*, and *hemorrhagic thrombocythemia*) is a clonal hematopoietic stem cell disorder associated with mutations in JAK2 (V617F), MPL, or CALR and manifested clinically by overproduction of platelets without a definable cause. ET has an incidence of 1–2/100,000 and a distinct female predominance. Canonical MPN driver mutations distinguish 90% of ET patients from the more common nonclonal, reactive forms of thrombocytosis (Table 103-6); mutation-negative ET patients may have either uncommon MPL mutations, JAK2 V617F expression limited to the platelets, or a hereditary form of thrombocytosis. Once considered a disease of the elderly and responsible for significant morbidity due to hemorrhage

TABLE 103-6 Causes of Thrombocytosis

Tissue inflammation: collagen vascular disease, inflammatory bowel disease	Hemorrhage
Malignancy	Iron-deficiency anemia
Infection	Surgery
Myeloproliferative disorders: polycythemia vera, primary myelofibrosis, essential thrombocytosis, chronic myelogenous leukemia	Rebound: Correction of vitamin B ₁₂ or folate deficiency, post-ethanol abuse
Myelodysplastic disorders: 5q-syndrome, idiopathic refractory sideroblastic anemia	Hemolysis
Postsplenectomy or hyposplenism	Familial: Thrombopoietin overproduction, JAK2 or MPL mutations

or thrombosis, it is now clear that ET can occur at any age in adults and often without symptoms or disturbances of hemostasis. There is an unexplained female predominance in contrast to PMF or the reactive forms of thrombocytosis where no sex difference exists. Because no specific clonal marker is available, clinical and laboratory criteria have been proposed to distinguish ET from other MPNs, which may also present with initially with isolated thrombocytosis but have differing prognoses and therapies (Table 103-6). These criteria are useful in identifying disorders such as CML, PV, PMF, or myelodysplasia, which can masquerade as ET. Furthermore, as with “idiopathic” erythrocytosis, nonclonal benign forms of thrombocytosis exist (e.g., hereditary overproduction of thrombopoietin and those with noncanonical JAK2 driver mutations) that are not widely recognized because we currently lack diagnostic assays. Approximately 50% of ET patients express JAK2 V617F, 30% CALR (both type 1 and type 2), and 8% MPL mutations. ET patients lacking a canonical MPN driver mutation usually have a benign prognosis.

ETIOLOGY

Megakaryocytopoiesis and platelet production depend on thrombopoietin and its receptor MPL. As in the case of early erythroid and myeloid progenitor cells, early megakaryocytic progenitors require the presence of interleukin 3 (IL-3) and stem cell factor for optimal proliferation in addition to thrombopoietin. Their subsequent terminal development is also enhanced by the chemokine stromal cell-derived factor 1 (SDF-1). Interestingly, terminal megakaryocyte maturation and platelet production do not require thrombopoietin.

Megakaryocytes are unique among hematopoietic progenitor cells because reduplication of their genome is endomitotic rather than mitotic and promoted by thrombopoietin. Unlike erythropoietin, thrombopoietin is produced primarily in the liver but has important functions in the bone marrow where it functions to maintain hematopoietic stem cells quiescent in their endosteal niches; once released from their niches, thrombopoietin promotes the proliferation of these cells in the sinusoidal niche. Like plasma erythropoietin and its target erythroblasts, an inverse correlation exists between the platelet count and plasma thrombopoietin. However, unlike erythropoietin, thrombopoietin is only constitutively produced and the plasma thrombopoietin level is controlled by the size of the platelet and megakaryocyte progenitor cell pools. Also, in contrast to erythropoietin, but like its myeloid counterparts, granulocyte and granulocyte-macrophage colony-stimulating factors, thrombopoietin not only enhances the proliferation of its target cells but also enhances the reactivity of their end-stage product, the platelet. Paradoxically, in the three MPNs, expression of the thrombopoietin receptor, MPL, is impaired and plasma thrombopoietin is increased despite the increased number of megakaryocytes and platelets.

The clonal nature of ET was established by analysis of glucose-6-phosphate dehydrogenase isoenzyme expression in patients hemizygous for this gene. Although thrombocytosis is its principal manifestation,

like the other MPNs, a hematopoietic stem cell is involved in ET. Furthermore, a number of families have been described in which ET was inherited, in one instance as an autosomal dominant trait. In addition to ET, PMF and PV have also been observed in such kindreds.

CLINICAL FEATURES

Clinically, ET is most often identified incidentally when a platelet count is obtained during the course of a routine medical evaluation. Occasionally, review of previous blood counts will reveal that an elevated platelet count was present but overlooked for many years. No symptoms or signs are specific for ET, but these patients can have hemorrhagic and thrombotic tendencies expressed as easy bruising for the former and microvascular occlusive events for the latter such as erythromelalgia, ocular migraine, or a TIA. Physical examination is generally unremarkable. Splenomegaly is indicative of another MPN, in particular PV, PMF, or CML.

Anemia is unusual, but a mild neutrophilic leukocytosis is not. The blood smear is most remarkable for the number of platelets present, some of which may be very large. The large mass of circulating platelets may prevent the accurate measurement of serum potassium due to release of platelet potassium upon blood clotting. This type of hyperkalemia is a test tube artifact and not associated with electrocardiographic abnormalities. Similarly, arterial oxygen measurements can be inaccurate unless thrombocytemic blood is collected on ice. The prothrombin and partial thromboplastin times are normal, whereas abnormalities of platelet function such as a prolonged bleeding time and impaired platelet aggregation can be present. However, despite much study, no platelet function abnormality is characteristic of ET, and no platelet function test predicts the risk of clinically significant bleeding or thrombosis.

The elevated platelet count may hinder marrow aspiration, but marrow biopsy usually reveals megakaryocyte hypertrophy and hyperplasia, as well as an overall increase in marrow cellularity. If marrow reticulin is increased, another diagnosis should be considered. The absence of stainable iron demands an explanation because iron deficiency alone can cause thrombocytosis, and absent marrow iron in the presence of marrow hypercellularity is a feature of PV.

Nonrandom cytogenetic abnormalities occur in ET but are uncommon, and no specific or consistent abnormality is notable, even those involving chromosomes 3 and 1, where the genes for thrombopoietin and its receptor, MPL, respectively, are located.

DIAGNOSIS

Thrombocytosis is encountered in a broad variety of clinical disorders (Table 103-6), in many of which inflammatory cytokine production is increased. The absolute level of the platelet count is not a useful diagnostic aid for distinguishing between benign and clonal causes of thrombocytosis. About 50% of ET patients express the *JAK2* V617F mutation. When *JAK2* V617F is absent, cytogenetic evaluation is mandatory to determine if the thrombocytosis is due to CML or a myelodysplastic disorder such as the 5q- syndrome or sideroblastic anemia. Because the *BCR-ABL* translocation can be present in the absence of the Ph chromosome, and because the *BCR-ABL* reverse transcriptase polymerase chain reaction is associated with false-positive results, fluorescence in situ hybridization (FISH) analysis for *BCR-ABL* is the preferred assay in patients with thrombocytosis in whom a cytogenetic study for the Ph chromosome is negative. *CALR* mutations (type 1 or type 2) are present in 30% and *MPL* mutations are present in 8% of ET patients who do not have a *JAK2* mutation. Anemia and ringed sideroblasts are not features of ET, but they are features of idiopathic refractory sideroblastic anemia, and in some of these patients, the thrombocytosis occurs in association with expression of *JAK2* V617F, *CALR*, or an *MPL* mutation. Significant splenomegaly should suggest the presence of another MPN, and in this setting, a red cell mass determination should be performed because splenomegaly can mask the presence of erythrocytosis. Importantly, what appears to be ET can evolve into PV (usually in women with *JAK2* V617F) or PMF (usually in men with type 1 *CALR* mutations) after a period of many years due to clonal evolution or succession. There is sufficient overlap of the

JAK2 V617F neutrophil allele burden between ET and PV that this cannot be used as a distinguishing diagnostic feature with the exception that, in ET, the quantitative *JAK2* V617F neutrophil allele is never greater than 50%; only a red cell mass and plasma volume determination can distinguish PV from ET, and importantly in this regard, 64% of *JAK2* V617F-positive ET patients in one study actually were found to have PV when red cell mass and plasma volume determinations were performed. Claims that ET and PV form a biological continuum are unfounded as these disorders have different gene expression profiles and different natural histories.

COMPLICATIONS

Perhaps no other condition in clinical medicine has caused otherwise astute physicians to intervene inappropriately more often than thrombocytosis, particularly if the platelet count is $>1 \times 10^6/\mu\text{L}$. It is commonly believed that a high platelet count causes thrombosis; however, no controlled clinical study has ever established this association, and in patients younger than age 60 years, the incidence of thrombosis was not greater in patients with thrombocytosis than in age-matched controls, and tobacco use appears to be the most important risk factor for thrombosis in ET patients.

To the contrary, very high platelet counts are associated primarily with hemorrhage due to acquired von Willebrand's disease. This is not meant to imply that an elevated platelet count cannot cause symptoms in an ET patient, but rather that the focus should be on the patient, not the platelet count. For example, some of the most dramatic neurologic problems in ET are migraine-related and respond only to lowering of the platelet count, whereas other symptoms such as erythromelalgia respond simply to platelet cyclooxygenase-1 inhibitors such as aspirin or ibuprofen, without a reduction in platelet number. Still others may represent an interaction between an atherosclerotic vascular system and a high platelet count, and others may have no relationship to the platelet count whatsoever. Recognition that PV can present with thrombocytosis alone as well as the discovery of previously unrecognized causes of hypercoagulability (Chaps. 116 and 117) make the older literature on the complications of thrombocytosis unreliable.

TREATMENT

Essential Thrombocytosis

Survival of ET patients is not different than the general population regardless of their driver mutation. An elevated platelet count in an asymptomatic patient without cardiovascular risk factors or tobacco use requires no therapy. Indeed, before any therapy is initiated in a patient with thrombocytosis, the cause of symptoms must be clearly identified as due to the elevated platelet count. When the platelet count rises above $1 \times 10^6/\mu\text{L}$, a substantial quantity of high-molecular-weight von Willebrand multimers are removed from the circulation and destroyed by the enlarged platelet mass, resulting in an acquired form of von Willebrand's disease. This can be identified by a reduction in ristocetin cofactor activity. In this situation, aspirin could promote hemorrhage. Bleeding in this situation is rarely spontaneous and usually responds to tranexamic acid or e-aminocaproic acid, which can be given prophylactically before and after elective surgery.

Plateletpheresis is at best a temporary and inefficient remedy that is rarely required. Importantly, ET patients treated with ^{32}P or alkylating agents are at risk of developing acute leukemia without any proof of benefit; combining either therapy with hydroxyurea increases this risk. If platelet reduction is deemed necessary on the basis of symptoms refractory to salicylates alone, pegylated IFN- α , the quinazoline derivative anagrelide, or hydroxyurea can be used to reduce the platelet count, but none of these is uniformly effective or without significant side effects. Hydroxyurea and aspirin were more effective than anagrelide and aspirin for prevention of TIA because hydroxyurea is a nitric oxide donor, but they were not more effective for the prevention of other types of arterial thrombosis and actually less effective for venous thrombosis. The risk of

gastrointestinal bleeding is also higher when aspirin is combined with anagrelide. Normalizing the platelet count does not prevent either arterial or venous thrombosis. Pegylated interferon can produce a complete molecular remission in some ET patients, but a role for it or ruxolitinib in ET management has not yet been established.

As more clinical experience is acquired, ET appears more benign than previously thought. Evolution to acute leukemia is more likely to be a consequence of therapy than of the disease itself. In managing patients with thrombocytosis, the physician's first obligation is to do no harm.

FURTHER READING

- ALVAREZ-LARRAN A et al: Antiplatelet therapy versus observation in low-risk essential thrombocythemia with CALR mutation. *Haematologica* 101:926, 2016.
- ASHER S et al: Current and future therapies for myelofibrosis. *Blood Rev* 42:100715, 2020.
- PASSAMONTI F et al: A clinical-molecular prognostic model to predict survival in patients with post polycythemia vera and post essential thrombocythemia myelofibrosis. *Leukemia* 31:2726, 2017.
- SPIVAK JL: How I treat polycythemia vera. *Blood* 134:341, 2019.

not yet fully understand why or how these secondary lesions occur. Patients with CHIP also have increased risk of cardiovascular mortality that is not fully explained. The link between these two seemingly unrelated issues (cardiovascular and hematologic malignancy) may lie in understanding the interactions between circulating clonally expanded blood cells and vascular endothelium. A "proinflammatory" state caused by clonal, infiltrating monocytes leads to accelerated atherosclerotic plaque development and altered cardiac remodeling. Similar phenomena may occur in the marrow and blood—altered relationships between hematopoietic stem cells with the marrow microenvironment along with altered immune surveillance. Both increase the likelihood that a clone may survive, acquire additional mutations, and then further expand eventually to leukemia. Whether early identification of CHIP in patients will provide therapeutic opportunities for patients remains to be seen. Certainly, modifying cardiovascular risk in patients with CHIP seems prudent, but development of mutation-directed therapies to eliminate problematic clones to prevent leukemia is likely to be more elusive.

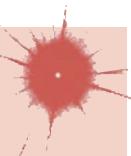
Genetic Predisposition Myeloid neoplasms typically occur sporadically in adults; inherited predisposition is rare. Yet, it is clear that myeloid neoplasms with germline predisposition represent an important and growing subset of disease. Germline mutations associated with increased risk of developing a myeloid neoplasm include *CEBPA*, *DDX41*, *RUNX1*, *ANKRD26*, *ETV6*, and *GATA2* (Table 104-1). Likewise, myeloid neoplasms with germline predisposition are a feature of several well-described clinical syndromes, including bone marrow failure disorders (e.g., Fanconi anemia, Shwachman-Diamond syndrome, Diamond-Blackfan anemia) and telomere biology disorders (e.g., dyskeratosis congenita). As new mutations and associations are added to a rapidly growing list, it is increasingly clear that genetic predisposition plays a larger role than has been previously understood.

Several genetic syndromes with somatic cell chromosome aneuploidy, such as Down syndrome with trisomy 21, are associated with an increased incidence of AML. Down syndrome-associated AML in young children (<4 years) is typically of the acute megakaryocytic subtype and is associated with mutation in the *GATA1* gene. Such

104

Acute Myeloid Leukemia

William Blum



INCIDENCE

Acute myeloid leukemia (AML) is a neoplasm characterized by infiltration of the blood, bone marrow, and other tissues by proliferative, clonal, poorly differentiated cells of the hematopoietic system. These leukemias comprise a spectrum of malignancies that untreated are uniformly fatal. In 2020, the estimated number of new AML cases in the United States was 19,940. AML is the diagnosis in 1.3% of all cancer cases and 31% of all new acute leukemias but causes 62% of leukemic deaths. AML is the most common acute leukemia in older patients, with a median age at diagnosis of 67 years. Long-term survival is infrequent; U.S. registry data report that only 27% of patients survive 5 years.

ETIOLOGY

Most cases of AML are idiopathic. Genetic predisposition, radiation, chemical/other occupational exposures, and drugs have been implicated in the development of AML, but AML cases with established etiology are relatively rare. No direct evidence suggests a viral etiology. Genome sequencing studies suggest that most cases of AML arise from a limited number of mutations that accumulate with advancing age. Indeed, genome sequencing is providing paradigm-shifting advances in our understanding of leukemogenesis. The Cancer Genome Atlas (TCGA) and other databases demonstrate that blood cells from up to 5–6% of normal individuals aged >70 years contain potentially "premalignant" mutations that are associated with clonal expansion.

Use of the term *premalignant* to describe these lesions is not precisely accurate; rather, these mutations represent clonal hematopoiesis of *indeterminate* potential (CHIP; sometimes called age-related clonal hematopoiesis). The genes most commonly altered include the epigenetic regulators *DNMT3A*, *TET2*, and *ASXL1*.

Study of CHIP is important because CHIP has relevance not just to blood cancer evolution but also other medical conditions. Clonal expansion driven by the acquisition of new mutations is associated with a 10-fold increase in risk for developing a hematologic malignancy (compared to matched patients without CHIP), but it is clear that additional "hits" must occur to drive toward leukemia. We do

TABLE 104-1 WHO 2016 Classification of Myeloid Neoplasms with Germline Predisposition

CLASSIFICATION^a

Myeloid neoplasms with germline predisposition without a preexisting disorder or organ dysfunction

Acute myeloid leukemia with germline *CEBPA* mutation

Myeloid neoplasms with germline *DDX41* mutation^b

Myeloid neoplasms with germline predisposition and preexisting platelet disorders

Myeloid neoplasms with germline *RUNX1* mutation^b

Myeloid neoplasms with germline *ANKRD26* mutation^b

Myeloid neoplasms with germline *ETV6* mutation^b

Myeloid neoplasms with germline predisposition and other organ dysfunction

Myeloid neoplasms with germline *GATA2* mutation

Myeloid neoplasms associated with bone marrow failure syndromes

Myeloid neoplasms associated with telomere biology disorders

Myeloid neoplasms associated with Noonan syndrome

Myeloid neoplasms associated with Down syndrome^b

^aRecognition of familial myeloid neoplasms requires that physicians take a thorough patient and family history to assess for typical signs and symptoms of known syndromes, including data on malignancies and previous bleeding episodes. Molecular genetic diagnostics is guided by a detailed patient and family history. Diagnostics should be performed in close collaboration with a genetic counselor; patients with a suspected heritable myeloid neoplasm, who test negative for known predisposition genes, should ideally be entered on a research study to facilitate new syndrome discovery. ^bLymphoid neoplasms also reported.

Source: Reproduced with permission from L Peterson et al: Myeloid neoplasms with germline predisposition, in *World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues*, 4th revised ed. Geneva, Switzerland: World Health Organization, 2017.

patients have excellent clinical outcomes but require dose modification of chemotherapy due to high treatment-related toxicities. Inherited diseases with defective DNA repair (e.g., Fanconi anemia, Bloom syndrome, and ataxia-telangiectasia) are also associated with AML. Each syndrome is associated with unique clinical features and atypical toxicities with chemotherapy, requiring expert care. Congenital neutropenia (Kostmann syndrome), due to mutations in the genes encoding the granulocyte colony-stimulating factor receptor and neutrophil elastase, is another disorder that may evolve into AML.

Chemical, Radiation, and Other Exposures Anticancer drugs are the leading cause of therapy-associated AML. Alkylating agent–associated leukemias occur on average 4–6 years after exposure, and affected individuals often have multilineage dysplasia and monosomy/aberrations in chromosomes 5 and 7. Topoisomerase II inhibitor–associated leukemias occur 1–3 years after exposure, and affected individuals often have AML with monocytic features and aberrations involving chromosome 11q23. Exposure to ionizing radiation, benzene, chloramphenicol, phenylbutazone, and other drugs can uncommonly result in bone marrow failure that may evolve into AML.

CLASSIFICATION

The current categorization of AML uses the World Health Organization (WHO) classification (Table 104-2), which defines biologically distinct groups based on cytogenetic and molecular abnormalities in addition to clinical features and light microscope morphology. Myeloid neoplasms with germline predisposition, as introduced above, are included as a new and important feature of this classification (Table 104-1).

TABLE 104-2 WHO 2016 Classification of Acute Myeloid Leukemia and Related Neoplasms

Acute myeloid leukemia (AML) with recurrent genetic abnormalities

- AML with t(8;21)(q22;q22); *RUNX1-RUNX1T1*
- AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*

Acute promyelocytic leukemia with PML-RARA

- AML with t(9;11)(p21.3;q23.3); *MLLT3-KMT2A*
- AML with t(6;9)(p23;q34.1); *DEK-NUP214*
- AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2, MECOM*
- AML (megakaryoblastic) with t(1;22)(p13.3;q13.3); *RBM15-MKL1*
- Provisional entity: AML with BCR-ABL1*
- AML with mutated *NPM1*
- AML with biallelic mutations of *CEBPA*
- Provisional entity: AML with mutated RUNX1*

AML with myelodysplasia-related changes

- Therapy-related myeloid neoplasms

AML, not otherwise specified (NOS)

- AML with minimal differentiation
- AML without maturation
- AML with maturation
- Acute myelomonocytic leukemia
- Acute monoblastic/monocytic leukemia
- Pure erythroid leukemia
- Acute megakaryoblastic leukemia
- Acute basophilic leukemia
- Acute panmyelosis with myelofibrosis

Myeloid sarcoma

Myeloid proliferations related to Down syndrome

- Transient abnormal myelopoiesis (TAM)
- Myeloid leukemia associated with Down syndrome

Note: Marrow blast count of ≥20% is required, except for AML with the recurrent genetic abnormalities t(15;17), t(8;21), inv(16), or t(16;16).

Source: Adapted from DA Arber et al: Acute myeloid leukaemia (AML) with recurrent genetic abnormalities, in *World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues*, 4th revised ed. Geneva, Switzerland: World Health Organization; 2017.

The WHO classification enables the identification of subsets of disease that may be treated differently (now or in the future) and enhances recognition of the molecular basis of disease from the time of diagnosis. Marrow (or blood) blast count of ≥20% is required to establish the diagnosis of AML, except for AML with the recurrent genetic abnormalities t(15;17), t(8;21), inv(16), or t(16;16).

Clinical Features Even with advances in molecular biology, recognizing clinical features remains important in understanding AML. For example, therapy-related AML is a distinct entity that develops following prior chemotherapy (e.g., alkylating agents, topoisomerase II inhibitors) or ionizing radiation. AML with myelodysplasia-related changes is recognized in part on morphology but also on a medical history of an antecedent myelodysplastic syndrome (MDS) or myelodysplastic/myeloproliferative neoplasm. These clinical features contribute to AML prognosis and are therefore in the WHO classification.



Genetic Findings Subtypes of AML are recognized due to the presence or absence of specific, recurrent cytogenetic, and/or genetic abnormalities. For example, the diagnosis of *acute promyelocytic leukemia* (APL) is based on the presence of either the t(15;17)(q22;q12) cytogenetic rearrangement or the *PML-RARA* fusion product of the translocation. Similarly, core binding factor (CBF) AML is designated based on the presence of t(8;21)(q22;q22), inv(16)(p13.1q22), or t(16;16)(p13.1;q22) or the respective fusion products *RUNX1-RUNX1T1* and *CBFB-MYH11*. Each of these three groups identifies patients with favorable clinical outcomes when appropriately treated.

Several cytogenetic or genetic AML subtypes often associate with a specific morphologic appearance, such as a complex karyotype (and/or mutation of *TP53*) and AML with myelodysplasia-related changes. Patients with such changes typically fare poorly with standard treatments. However, only one abnormality is invariably associated with specific morphologic features: t(15;17)(q22;q12) or the molecular fusion *PML-RARA* with APL. Other cytogenetic and genetic findings may be commonly, but not always, associated with a morphologic description, highlighting the necessity of genetic and cytogenetic testing for precise diagnosis. Several chromosomal abnormalities often associate primarily with one morphologic/immunophenotypic group. Examples include inv(16)(p13.1q22) with AML with abnormal bone marrow eosinophils; t(8;21)(q22;q22) with slender Auer rods, expression of CD19, and increased normal eosinophils; and t(9;11)(p22;q23) and other translocations involving 11q23 with monocytic features. Mutation of nucleophosmin (nucleolar phosphoprotein B23, numatrin, *NPM1*), especially when co-occurring with mutation of fms-related tyrosine kinase 3 (*FLT3*), often presents with “cup-shaped” nuclear morphology. Recurring chromosomal abnormalities in AML may also be loosely associated with specific clinical characteristics. More commonly associated with younger age are t(8;21) and t(15;17), and with older age, del(5q), del(7q), and mutated *TP53*. Myeloid sarcomas are associated with t(8;21); disseminated intravascular coagulation (DIC) is associated with t(15;17). 11q23 aberrations and monocytic leukemia are associated with extramedullary sites of involvement at presentation, especially gingival hypertrophy. High leukocyte count is commonly observed with *NPM1* or *FLT3* mutation.

The WHO classification also incorporates molecular abnormalities by recognizing fusion genes or specific genetic mutations with a role in leukemogenesis. As a classic example, t(15;17) results in the fusion gene *PML-RARA* that encodes a chimeric protein, promyelocytic leukemia (Pml)–retinoic acid receptor α (Rara), which is formed by the fusion of the retinoic acid receptor α (*RARA*) gene from chromosome 17 and the promyelocytic leukemia (*PML*) gene from chromosome 15. Unique clinical therapy with retinoic acid and arsenic trioxide has revolutionized the care of APL patients (see “Treatment of Acute Promyelocytic Leukemia” section). Similar examples of molecular subtypes included in the category of AML with recurrent genetic abnormalities are those characterized by the leukemogenic fusion genes *RUNX1-RUNX1T1* and *CBFB-MYH11* and the so-called CBF AML subtypes noted cytogenetically as t(8;21), inv(16), or t(16;16). Additional fusions

are *MLLT3-KMT2A* and *DEK-NUP214*, resulting from t(9;11) and t(6;9)(p23;q34), respectively, among others.

The WHO classification of AML continues to expand as knowledge of specific genetic or cytogenetic aberrations grows. Several AML subtypes are defined by the presence of genetic mutations rather than chromosomal aberrations. For example, *AML with mutated NPM1* and *AML with biallelic mutated CEBPA*, respectively, are associated with more favorable clinical outcome, though the presence of coexisting mutation in *FLT3* affects *NPM1* prognostic impact. Activating mutations of *FLT3* are present in ~30% of adult AML patients, primarily due to internal tandem duplications (ITDs) in the juxtamembrane domain that have negative prognostic impact. In contrast, point mutations of the activating loop of the kinase (called tyrosine kinase domain [TKD] mutations) have uncertain prognostic impact. Aberrant activation of the *FLT3*-encoded protein provides increased proliferation and antiapoptotic signals to the myeloid progenitor cell. *FLT3-ITD*, the more common of the *FLT3* mutations, occurs preferentially in patients with cytogenetically normal AML (CN-AML). The importance of identifying *FLT3-ITD* at diagnosis relates to the fact that it is not only useful as a prognosticator but also may predict response to specific treatment such as a tyrosine kinase inhibitor (TKI). Several TKIs targeting *FLT3* are either approved for AML (e.g., midostaurin, only in first-line therapy in combination with chemotherapy; gilteritinib, in relapse as monotherapy) or currently in clinical investigation (e.g., quizartinib, crenolanib, sorafenib, and others). The *FLT3* allelic ratio (of the number of mutated alleles to wild-type alleles) provides information beyond the mere presence or absence of the mutation. Several mutational scenarios, such as one mutated gene and one wild-type gene or one mutated gene with no (deleted) wild-type gene, and the ratio of malignant to nonmalignant cells in the sample affect the ratio. The allelic ratio affects the prognostic impact of the *FLT3-ITD* mutation; patients with *FLT3-ITD* “low” allelic ratio (<0.5) fare better. Accordingly, mutated *NPM1* without *FLT3-ITD* or with *FLT3-ITD*^{low} is viewed as favorable risk by the European LeukemiaNet (ELN) risk stratification schema (Table 104-3). Conversely, *FLT3-ITD*^{high} has an adverse prognostic impact; patients with both mutated *NPM1* and *FLT3-ITD* with an allelic ratio >0.5 are intermediate risk by ELN stratification. Involving a different tyrosine kinase, AML with *BCR-ABL1* fusion is a new WHO provisional entity to recognize rare cases that may benefit from *BCR-ABL* TKI therapy (Table 104-2).

Immunophenotypic Findings The immunophenotype of human leukemia cells can be studied by multiparameter flow cytometry after the cells are labeled with monoclonal antibodies to cell-surface antigens. This can be important in quickly distinguishing AML from acute lymphoblastic leukemia and for identifying some subtypes of AML. For example, AML with minimal differentiation, characterized by immature morphology and no lineage-specific cytochemical reactions, may be diagnosed by flow-cytometric demonstration of the myeloid-specific antigens cluster designation (CD) 13 and/or 117. Similarly, acute megakaryoblastic leukemia can often be diagnosed only by expression of the platelet-specific antigens CD41 and/or CD61. Although flow cytometry is widely used, and in some cases essential for the diagnosis of AML, it has only a supportive role in establishing the different subtypes of AML through the WHO classification. Increasingly, multiparameter flow cytometry is used for the measurement of measurable residual disease (MRD) after remission is achieved.

■ PROGNOSTIC FACTORS

Several factors predict outcome of AML patients treated with chemotherapy; they should be used for risk stratification and treatment guidance.

Chromosome and molecular investigations performed at diagnosis currently provide the most important prognostic information. WHO has categorized patients as having favorable, intermediate, or adverse risk based on the presence of structural and/or numerical chromosomal or genetic aberrations. Patients with t(15;17) have a very good prognosis (~85% cured), and those with t(8;21) and inv(16) have a good prognosis (~55% cured), whereas those with no cytogenetic

TABLE 104-3 2017 European LeukemiaNet Risk Stratification by Genetics for Acute Myeloid Leukemia (AML)^a

RISK CATEGORY ^b	GENETIC ABNORMALITY
Favorable	t(8;21)(q22;q22); <i>RUNX1-RUNX1T1</i> inv(16)(p13.1;q22) or t(16;16)(p13.1;q22); <i>CBFB-MYH11</i> Mutated <i>NPM1</i> without <i>FLT3-ITD</i> or with <i>FLT3-ITD</i> ^{low(c)} Biallelic mutated <i>CEBPA</i>
Intermediate	Mutated <i>NPM1</i> and <i>FLT3-ITD</i> ^{high(c)} Wild-type <i>NPM1</i> without <i>FLT3-ITD</i> or with <i>FLT3-ITD</i> ^{low(c)} (w/o adverse-risk genetic lesions) t(9;11)(p21.3;q23.3); <i>MLLT3-KMT2A</i> ^d Cytogenetic abnormalities not classified as favorable or adverse
Adverse	t(6;9)(p23;q34.1); <i>DEK-NUP214</i> t(v;11q23.3); <i>KMT2A</i> rearranged t(9;22)(q34.1;q11.2); <i>BCR-ABL1</i> inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); <i>GATA2</i> , <i>MECOM(EVI1)</i> –5 or del(5q); –7; –17/abn(17p) Complex karyotype, ^e monosomal karyotype ^f Wild type <i>NPM1</i> and <i>FLT3-ITD</i> ^{high(c)} Mutated <i>RUNX1</i> ^g Mutated <i>ASXL1</i> ^g Mutated <i>TP53</i> ^h

^aThis table excludes acute promyelocytic leukemia. Frequencies, response rates, and outcome measures should be reported by risk category and, if sufficient numbers are available, by specific genetic lesions indicated. ^bPrognostic impact of a marker is treatment-dependent and may change with new therapies. ^cLow, low allelic ratio (<0.5); high, high allelic ratio (>0.5); semiquantitative assessment of *FLT3-ITD* allelic ratio (using DNA fragment analysis) is determined as ratio of the area under the curve (AUC) “*FLT3-ITD*” divided by AUC “*FLT3-wild type*”; recent studies indicate that acute myeloid leukemia with *NPM1* mutation and *FLT3-ITD* low allelic ratio may also have a more favorable prognosis and patients should not routinely be assigned to allogeneic hematopoietic cell transplantation. ^dThe presence of t(9;11)(p21.3;q23.3) takes precedence over rare, concurrent adverse-risk gene mutations. ^eThree or more unrelated chromosome abnormalities in the absence of one of the World Health Organization–designated recurring translocations or inversions, i.e., t(8;21), inv(16) or t(16;16), t(9;11), t(v;11q23.3), t(6;9), inv(3), or t(3;3); AML with *BCR-ABL1*. ^fDefined by the presence of one single monosomy (excluding loss of X or Y) in association with at least one additional monosomy or structural chromosome abnormality (excluding core binding factor AML). ^gThese markers should not be used as an adverse prognostic marker if they co-occur with favorable-risk AML subtypes. ^h*TP53* mutations are significantly associated with AML with complex and monosomal karyotype.

Source: Republished with permission of American Society of Hematology, from Diagnosis and management of acute myeloid leukemia in adults: 2017 recommendations from an international expert panel, Döhner H et al. 129:424, 2017; permission conveyed through Copyright Clearance Center, Inc.

abnormality have an intermediate outcome risk (~40% cured). Patients with a *TP53* mutation, complex karyotype, t(6;9), inv(3), or –7 have a very poor prognosis. Another cytogenetic subgroup, the monosomal karyotype, has been suggested to adversely influence the outcome of AML patients other than those with t(15;17), t(8;21), or inv(16) or t(16;16). The monosomal karyotype subgroup is defined by the presence of at least two autosomal monosomies (loss of chromosomes other than Y or X) or a single autosomal monosomy with additional structural abnormalities.

For patients lacking prognostic cytogenetic abnormalities, i.e., those with CN-AML, testing for several mutated genes can help to risk-stratify. In addition to *NPM1* mutation and *FLT3-ITD* as described above, biallelic *CEBPA* mutations have prognostic value. Such mutations predict favorable outcome. Given the proven prognostic importance of *NPM1*, *CEBPA*, and *FLT3*, molecular assessment of these genes at diagnosis has been incorporated into AML management guidelines by the National Comprehensive Cancer Network (NCCN) and the ELN. The same markers help to define genetic groups in the ELN standardized reporting system, which is based on both cytogenetic and molecular abnormalities and is used for comparing clinical features/treatment response among subsets of patients reported across different clinical studies (Table 104-3). These genetic groups should be used for risk stratification and treatment guidance.

TABLE 104-4 Molecular Prognostic Markers in AML^a

GENE SYMBOL	GENE LOCATION	PROGNOSTIC IMPACT
Genes Included in the WHO Classification and ELN Reporting System		
<i>NPM1</i> mutations	5q35.1	Favorable
<i>CEBPA</i> mutations	19q13.1	Favorable
<i>FLT3</i> -ITD	13q12	Depends on allelic ratio and <i>NPM1</i> mutational status
Genes Encoding Receptor Tyrosine Kinases		
<i>KIT</i> mutation	4q12	Adverse
<i>FLT3</i> -TKD	13q12	Unclear
Genes Encoding Transcription Factors		
<i>RUNX1</i> mutations	21q22.12	Adverse
<i>WT1</i> mutations	11p13	Adverse
Genes Encoding Epigenetic Modifiers		
<i>ASXL1</i> mutations	20q11.21	Adverse
<i>DNMT3A</i> mutations	2p23.3	Adverse
<i>IDH</i> mutations (<i>IDH1</i> and <i>IDH2</i>)	2q34 & 15q26.1	Adverse
<i>KMT2A</i> -PTD	11q23	Adverse
<i>TET2</i> mutations	4q24	Adverse
Deregulated Genes		
<i>BAALC</i> overexpression	8q22.3	Adverse
<i>ERG</i> overexpression	21q22.3	Adverse
<i>MN1</i> overexpression	22q12.1	Adverse
<i>EVI1</i> overexpression	3q26.2	Adverse
Deregulated MicroRNAs		
<i>miR-155</i> overexpression	21q21.3	Adverse
<i>miR-3151</i> overexpression	8q22.3	Adverse
<i>miR-181a</i> overexpression	1q32.1 and 9q33.3	Favorable

^aThis table excludes acute promyelocytic leukemia.

Abbreviations: AML, acute myeloid leukemia; ELN, European LeukemiaNet; ITD, internal tandem duplication; PTD, partial tandem duplication; TKD, tyrosine kinase domain; WHO, World Health Organization.

In addition to *NPM1*, *CEBPA*, *FLT3*, and *TP53* mutations, molecular aberrations in other genes may be routinely used for prognostication (Table 104-4). Among these mutated genes are those encoding receptor tyrosine kinases (*KIT*), transcription factors (*RUNX1* and *WT1*), and epigenetic modifiers (*ASXL1*, *DNMT3A*, isocitrate dehydrogenase 1 [*IDH1*], *IDH2*, *KMT2A* [also known as *MLL*], and *TET2*). Although *KIT* mutations are almost exclusively present in CBF AML and impact adversely the outcome, the remaining markers have been reported primarily in CN-AML. Mutations of *ASXL1* and *RUNX1* are associated with adverse outcome, independent of other prognostic factors. However, for some of these mutations, data remain unclear on the prognostic impact due to conflicting reports (e.g., *TET2*, *IDH1*, *IDH2*). Increasingly, novel drugs that inhibit/modulate aberrant pathways activated by some of these genes (especially *FLT3*, *IDH1*, and *IDH2*) have been remarkably effective in subsets of disease, leading to U.S. Food and Drug Administration approvals (see section on treatment of AML).

In addition to gene mutations, deregulation of the expression levels of coding genes and of short noncoding RNAs (microRNAs) also provides prognostic information (Table 104-4). Overexpression of genes such as *BAALC*, *ERG*, *MN1*, and *MDS1* and *EVI1* complex locus (*MECOM*; also known as *EVI1*) predict poor outcome, especially in CN-AML. Similarly, deregulated expression levels of microRNAs, naturally occurring noncoding RNAs that regulate the expression of proteins via degradation or translational inhibition of their target coding RNAs, have also been associated with prognosis in AML. Overexpression of *miR-155* and *miR-3151* predicts unfavorable outcome in CN-AML, whereas overexpression of *miR-181a* predicts favorable outcome both in CN-AML and cytogenetically abnormal AML.

Because prognostic molecular markers in AML are not mutually exclusive and often occur concurrently (>80% patients have at least two or more prognostic gene mutations), the likelihood that distinct marker combinations may be more informative than single markers is increasingly clear.

Epigenetic changes (e.g., DNA methylation and/or posttranslational histone modification) and microRNAs are often involved in deregulation of genes involved in hematopoiesis, contribute to leukemogenesis, and may associate with the previously discussed prognostic gene mutations. These changes have been shown to provide biologic insights into leukemogenic mechanisms and provide independent prognostic information. Therapeutic progress based on advances in understanding the role of epigenetic changes in AML is currently unfolding. For example, in patients with mutations of *IDH1* or *IDH2*, novel active enzymes produced from these respective mutations hijack the citric acid cycle, leading to production of a novel “oncometabolite,” 2-hydroxyglutarate, which disrupts a myriad of epigenetic processes. Pharmacologic inhibition of these aberrant enzymes can reverse these leukemogenic activities.

In addition to cytogenetic and molecular aberrations, several other factors are associated with outcome in AML. Age at diagnosis is one of the most important risk factors. Advancing age is associated with a poor prognosis for two reasons: (1) its influence on the ability to survive induction therapy due to coexisting medical comorbidities, and (2) with each successive decade of age, a greater proportion of patients have intrinsically more resistant disease. A prolonged symptomatic interval with cytopenias preceding AML diagnosis or a history of antecedent hematologic disorders including MDS or myeloproliferative neoplasms is often found in older patients. Cytopenia is a clinical feature associated with a lower complete remission (CR) rate and shorter survival time. The CR rate is lower in patients who have had anemia, leukopenia, and/or thrombocytopenia for >3 months before the diagnosis of AML, when compared to those without such a history. Responsiveness to chemotherapy declines as the duration of the antecedent disorder increases. Likewise, AML developing after treatment with cytotoxic agents for other malignancies is usually difficult to treat successfully. In addition, older patients less frequently harbor favorable cytogenetic abnormalities (i.e., t[8;21], inv[16], and t[16;16]) and more frequently harbor adverse cytogenetic (e.g., complex and monosomal karyotypes) and/or molecular (e.g., *ASXL1*, *TP53*) abnormalities.

Other factors independently associated with worse outcome are a poor performance status that influences ability to survive induction therapy and a high presenting leukocyte count that in some series is an adverse prognostic factor for attaining a CR. Among patients with hyperleukocytosis (>100,000/ μ L), early central nervous system bleeding and pulmonary leukostasis contribute to poor outcomes.

Following administration of therapy, achievement of CR is associated with better outcome and longer survival. CR is defined after examination of both blood and bone marrow and essentially represents eradication of detectable leukemia and restoration of normal hematopoiesis. The blood neutrophil count must be $\geq 1000/\mu$ L and the platelet count $\geq 100,000/\mu$ L. Hemoglobin concentration is not considered in determining CR. Circulating blasts should be absent. Although rare blasts may be detected in the blood during marrow regeneration, they should disappear on successive studies. At CR, the bone marrow should contain <5% blasts, and Auer rods should be absent. Extramedullary leukemia should not be present.

CLINICAL PRESENTATION

Symptoms Patients with AML usually present with nonspecific symptoms that begin gradually, or abruptly, and are the consequence of anemia, leukocytosis, leukopenia/leukocyte dysfunction, or thrombocytopenia. Nearly half have symptoms for ≤ 3 months before the leukemia is diagnosed.

Fatigue is a frequent first symptom among AML patients. Anorexia and weight loss are common. Fever with or without an identifiable infection is the initial symptom in ~10% of patients. Signs of abnormal hemostasis (bleeding, easy bruising) are common. Bone pain,

lymphadenopathy, nonspecific cough, headache, or diaphoresis may also occur.

Rarely, patients may present with symptoms from a myeloid sarcoma (a tumor mass consisting of myeloid blasts occurring at anatomic sites other than bone marrow). Sites involved are most commonly the skin, lymph node, gastrointestinal tract, soft tissue, and testis. This rare presentation, often characterized by chromosome aberrations (e.g., monosomy 7, trisomy 8, 11q23 rearrangement, inv[16], trisomy 4, t[8;21]), may precede or coincide with blood and/or marrow involvement by AML. Patients who present with isolated myeloid sarcoma typically develop blood and/or marrow involvement quickly thereafter and cannot be cured with local therapy (radiation or surgery) alone.

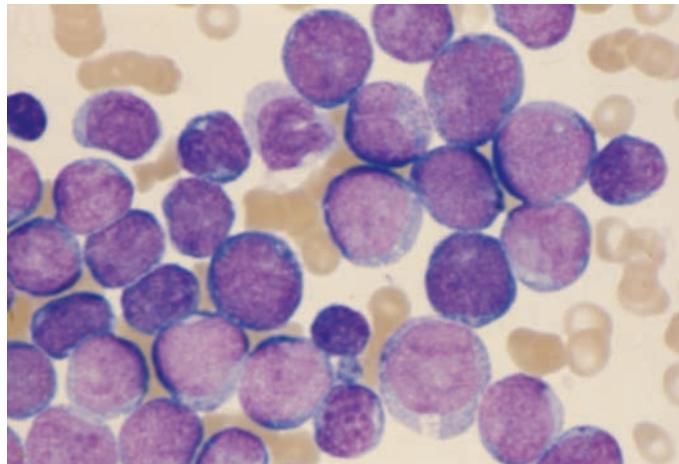
Physical Findings Fever, infection, and hemorrhage are often found at the time of diagnosis; splenomegaly, hepatomegaly, lymphadenopathy, and “bone pain” may also be present less commonly. Hemorrhagic complications are most commonly and, classically, found in APL. APL patients often present with DIC-associated minor hemorrhage but may have significant gastrointestinal bleeding, intrapulmonary hemorrhage, or intracranial hemorrhage. Likewise, thrombosis is another less frequent but well recognized clinical feature of DIC in APL. Bleeding associated with coagulopathy may also occur in monocytic AML and with extreme degrees of leukocytosis or thrombocytopenia in other morphologic subtypes. Retinal hemorrhages are detected in 15% of patients. Infiltration of the gingiva, skin, soft tissues, or meninges with leukemic blasts at diagnosis is characteristic of the monocytic subtypes and those with 11q23 chromosomal abnormalities.

Hematologic Findings Anemia is usually present at diagnosis, although it is not typically severe. The anemia is usually normocytic normochromic. Decreased erythropoiesis in the setting of AML often results in a reduced reticulocyte count, and red blood cell (RBC) survival is decreased by accelerated destruction. Active blood loss may rarely contribute to the anemia.

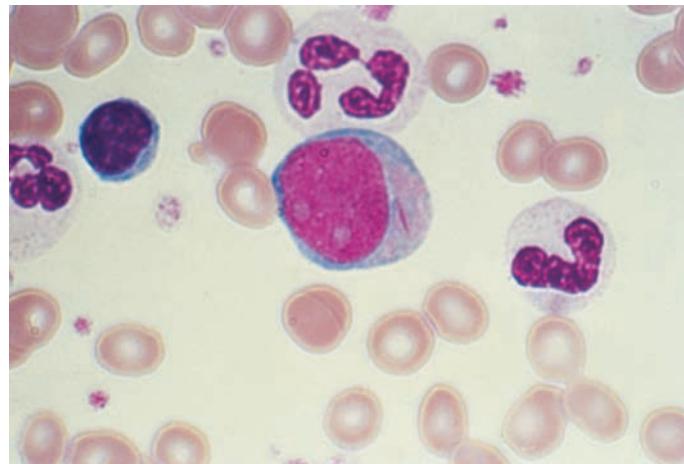
The median presenting leukocyte count is ~15,000/ μ L. Lower presenting leukocyte counts are more typical of older patients and those with antecedent hematologic disorders. Between 25 and 40% of patients have counts <5000/ μ L, and 20% have counts >100,000/ μ L. Fewer than 5% have no detectable leukemic cells in the blood. In AML, the cytoplasm often contains primary (nonspecific) granules, and the nucleus shows fine, lacy chromatin with one or more nucleoli characteristic of immature cells. Abnormal rod-shaped granules called Auer rods are not uniformly present, but when they are, AML is virtually certain (Fig. 104-1).

Platelet counts <100,000/ μ L are found at diagnosis in ~75% of patients, and ~25% have counts <25,000/ μ L. Both morphologic and functional platelet abnormalities can be observed, including large and bizarre shapes with abnormal granulation and inability of platelets to aggregate or adhere normally to one another.

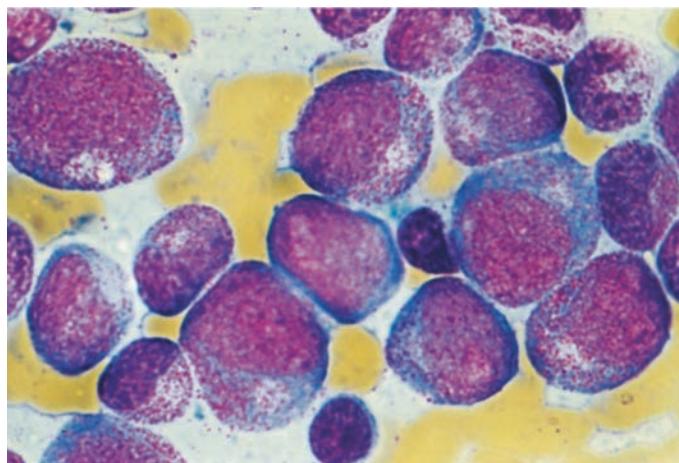
Pretreatment Evaluation Once the diagnosis of AML is suspected, thorough evaluation and initiation of appropriate therapy should follow. In addition to clarifying the subtype of leukemia, initial studies should evaluate the overall functional integrity of the major organ systems, including the cardiovascular, pulmonary, hepatic, and



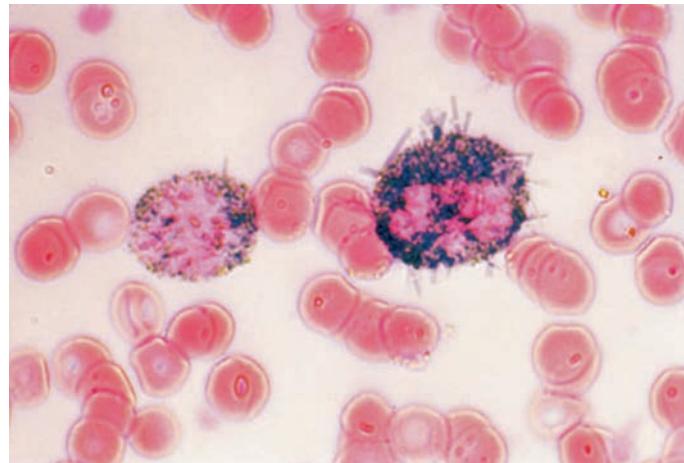
A



B



C



D

FIGURE 104-1 Morphology of acute myeloid leukemia (AML) cells. **A.** Uniform population of primitive myeloblasts with immature chromatin, nucleoli in some cells, and primary cytoplasmic granules. **B.** Leukemic myeloblast containing an Auer rod. **C.** Promyelocytic leukemia cells with prominent cytoplasmic primary granules. **D.** Peroxidase stain shows dark blue color characteristic of peroxidase in granules in AML.

TABLE 104-5 Initial Diagnostic Evaluation and Management of Adult Patients with AML**History**

- Increasing fatigue or decreased exercise tolerance (anemia)
- Excess bleeding or bleeding from unusual sites (DIC, thrombocytopenia)
- Fever or recurrent infections (neutropenia)
- Headache, vision changes, nonfocal neurologic abnormalities (CNS leukemia or bleed)
- Early satiety (splenomegaly)
- Family history of AML (Fanconi, Bloom, or Kostmann syndromes or ataxia-telangiectasia)
- History of cancer (exposure to alkylating agents, radiation, topoisomerase II inhibitors)
- Occupational exposures (radiation, benzene, petroleum products, paint, smoking, pesticides)

Physical Examination

- Performance status (prognostic factor)
- Ecchymosis and oozing from IV sites (DIC, possible acute promyelocytic leukemia)
- Fever and tachycardia (signs of infection)
- Papilledema, retinal infiltrates, cranial nerve abnormalities (CNS leukemia)
- Poor dentition, dental abscesses
- Gum hypertrophy (leukemic infiltration, most common in monocytic leukemia)
- Skin infiltration or nodules (leukemia infiltration, most common in monocytic leukemia)
- Lymphadenopathy, splenomegaly, hepatomegaly
- Back pain, lower extremity weakness (spinal granulocytic sarcoma, most likely in t[8;21] patients)

Laboratory and Radiologic Studies

- CBC with manual differential cell count
- Chemistry tests (electrolytes, creatinine, BUN, calcium, phosphorus, uric acid, hepatic enzymes, bilirubin, LDH, amylase, lipase)
- Clotting studies (prothrombin time, partial thromboplastin time, fibrinogen, D-dimer)
- Viral serologies (CMV, HSV-1, varicella-zoster)
- RBC type and screen
- HLA typing for potential allogeneic HCT
- Bone marrow aspirate and biopsy (morphology, cytogenetics, flow cytometry, molecular studies for *NPM1* and *CEBPA* mutations and *FLT3-ITD*)
- Cryopreservation of viable leukemia cells
- Myocardial function (echocardiogram or MUGA scan)
- PA and lateral chest radiograph
- Placement of central venous access device

Interventions for Specific Patients

- Dental evaluation (for those with poor dentition)
- Lumbar puncture (for those with symptoms of CNS involvement)
- Screening spine MRI (for patients with back pain, lower extremity weakness, paresthesias)
- Social work referral for patient and family psychosocial support

Counseling for All Patients

Provide patients with information regarding their disease and genetic risks, sperm banking or menstrual suppression, financial counseling, and support group contact

Abbreviations: AML, acute myeloid leukemia; BUN, blood urea nitrogen; CBC, complete blood count; CMV, cytomegalovirus; CNS, central nervous system; DIC, disseminated intravascular coagulation; HLA, human leukocyte antigen; HCT, hematopoietic stem cell transplantation; HSV, herpes simplex virus; IV, intravenous; LDH, lactate dehydrogenase; MRI, magnetic resonance imaging; MUGA, multigated acquisition; PA, posteroanterior; RBC, red blood (cell) count.

renal systems (**Table 104-5**). Factors that have prognostic significance, either for achieving CR or for CR duration, should also be assessed before initiating treatment including cytogenetics and molecular markers. Leukemic cells should be obtained from all patients and

cryopreserved for future investigational testing as well as potential future use as new diagnostics and therapeutics become available. All patients should be evaluated for infection. During the ongoing global pandemic, testing for the presence of the novel coronavirus, SARS-CoV2, is recommended before initiation of chemotherapy.

Most patients are anemic and thrombocytopenic at presentation. Replacement of the appropriate blood components, if necessary, should begin promptly. Because qualitative platelet dysfunction or the presence of an infection may increase the likelihood of bleeding, evidence of hemorrhage justifies the immediate use of platelet transfusion, even if the platelet count is only moderately decreased.

About 50% of patients have a mild to moderate elevation of serum uric acid at presentation. Only 10% have marked elevations, but renal precipitation of uric acid and the nephropathy that may result is a serious but uncommon complication. The initiation of chemotherapy may aggravate hyperuricemia, and patients are usually started immediately on allopurinol and hydration at diagnosis. Rasburicase (recombinant uric oxidase) is also useful for treating uric acid nephropathy and often can normalize the serum uric acid level within hours with a single dose of treatment, although its expense suggests that limiting its use to patients with severe hyperuricemia and/or kidney injury may be prudent. The presence of high concentrations of lysozyme, a marker for monocytic differentiation, may be etiologic in renal tubular dysfunction for a minority of patients.

TREATMENT**Acute Myeloid Leukemia**

Treatment of the newly diagnosed patient with AML is usually divided into two phases, induction and postremission management (consolidation) (**Fig. 104-2**). The initial goal is to induce CR. Once CR is obtained, further therapy must be given to prolong survival and achieve cure. The initial induction treatment and subsequent postremission therapy are chosen based on the patient's age, overall fitness, and cytogenetic/molecular risk. Intensive therapy with cytarabine and anthracycline in younger patients (<60 years) increases the cure rate of AML. In older patients, the benefit of intensive therapy is controversial in all but favorable-risk patients; novel approaches for selecting patients predicted to be responsive to treatment and new therapies are being pursued. Additional options for therapy have emerged for older AML patients such as the addition of the BCL2 antagonist venetoclax to one of several low-intensity chemotherapies. Likewise, novel oral drugs targeting IDH1 or IDH2, alone or in combination with low-intensity chemotherapy, may be considered as initial therapy for older patients who have mutations in those respective pathways.

INDUCTION CHEMOTHERAPY

The most commonly used induction regimens (for patients other than those with APL) consist of combination chemotherapy with cytarabine and an anthracycline (e.g., daunorubicin, idarubicin). Cytarabine is a cell cycle S-phase-specific antimetabolite that becomes phosphorylated intracellularly to an active triphosphate form that interferes with DNA synthesis. Anthracyclines are DNA intercalators. Their primary mode of action is thought to be inhibition of topoisomerase II, leading to DNA breaks.

In adults, cytarabine used at standard dose (100–200 mg/m²) is administered as a continuous intravenous infusion for 7 days. With cytarabine, anthracycline therapy generally consists of daunorubicin (60–90 mg/m²) or idarubicin (12 mg/m²) intravenously on days 1, 2, and 3 (the 7 and 3 regimen). Other agents can be added (e.g., gemtuzumab ozogamicin) when 60 mg/m² of daunorubicin is used. With the 7 and 3 regimen, it is now clearly established that 45 mg/m² dosing of daunorubicin results in inferior outcomes; patients should receive higher doses as described. Patients failing remission after one induction are offered reinduction with the same (or slightly modified) therapy. The CD33-targeting immunoconjugate gemtuzumab ozogamicin may be added to induction therapy for subsets of patients, especially those with CBF AML.

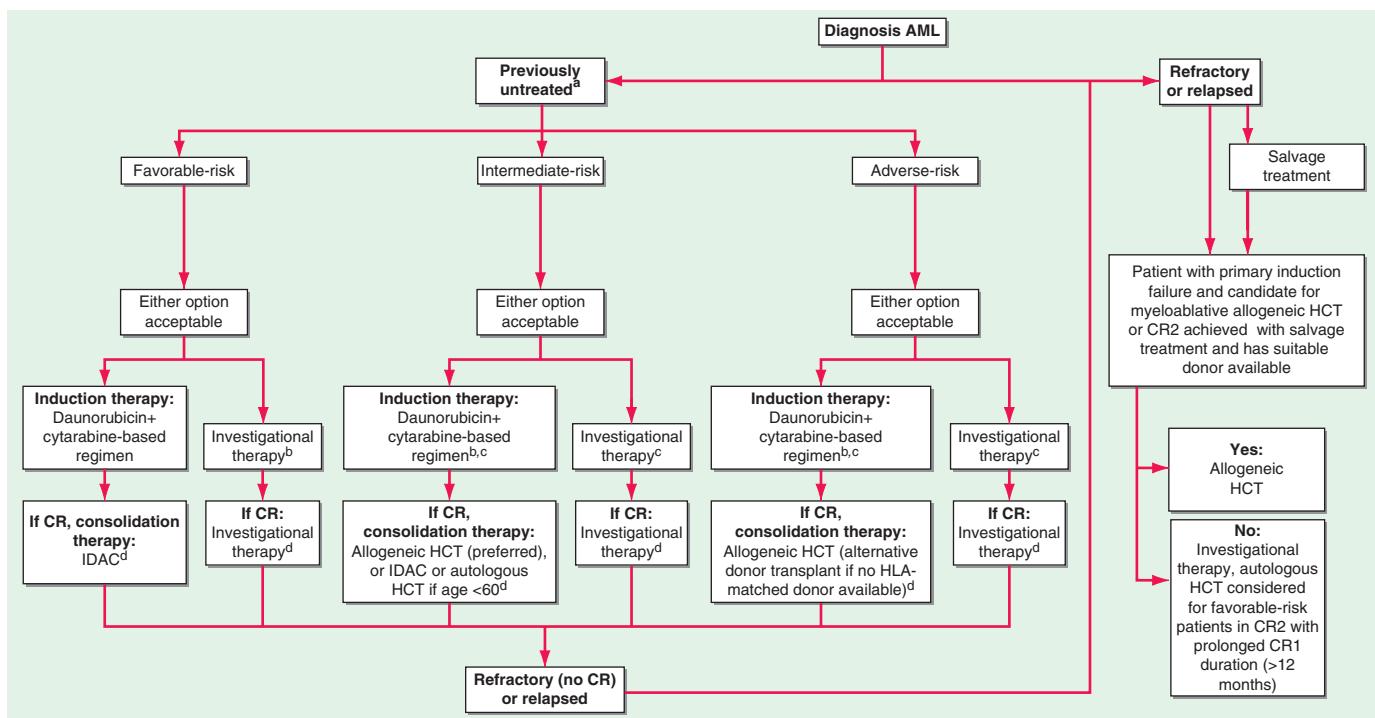


FIGURE 104-2 Algorithm for the therapy of newly diagnosed acute myeloid leukemia (AML). ^aRisk stratification according to the European LeukemiaNet (see Table 104-3). ^bYounger patients (<60–65 years) should routinely be offered investigational therapy on a backbone of standard chemotherapy for induction and consolidation. ^cOlder patients, especially those >65 years or with adverse risk disease, or those who are unfit for intensive anthracycline + cytarabine regimens, may be considered for investigational therapy alone or in combination with lower intensity chemotherapy (azacitidine, decitabine, cytarabine), or lower intensity chemotherapy in combination with venetoclax. ^dInvestigational therapy as maintenance should be considered if available (after consolidation for younger patients and older patients with favorable-risk disease, and for all other older patients after induction).

Allogeneic hematopoietic cell transplantation (HCT) is a consideration for all eligible patients in first complete remission (CR) with non–favorable-risk disease and highly recommended for older patients (60–75 years) and those with adverse risk.

For all forms of AML in fit patients, except acute promyelocytic leukemia (APL), standard induction therapy includes a regimen based on a 7-day continuous infusion of cytarabine (100–200 mg/m²/d) and a 3-day course of daunorubicin (60–90 mg/m²/d) with or without additional drugs. Idarubicin (12 mg/m²/d) can be used in place of daunorubicin (not shown). The value of postremission/consolidation therapy for older patients (>60 years) who do not have favorable-risk disease is uncertain. Patients who achieve CR undergo postremission consolidation therapy, including sequential courses of intermediate-dose cytarabine, allogeneic HCT, autologous HCT, or novel therapies, based on their predicted risk of relapse (i.e., risk-stratified therapy). Patients receiving induction of lower intensity chemotherapy with venetoclax (or investigational therapy) typically receive repetitive cycles of same on an attenuated schedule, if necessary due to myelotoxicity, after achieving remission. Patients with APL (see text for treatment) usually receive tretinoin and arsenic trioxide–based regimens with or without anthracycline-based chemotherapy and possibly maintenance with tretinoin. HLA, human leukocyte antigen; IDAC, intermediate-dose cytarabine.

In older patients (age ≥60–65 years), the outcome with conventional intensive therapy is generally poor due to a higher frequency of resistant disease and increased rate of treatment-related mortality. This is especially true in patients with prior hematologic disorders (MDS or myeloproliferative neoplasms), therapy-related AML, or cytogenetic and genetic abnormalities that adversely influence clinical outcome. Patients still fare far better with treatment than with supportive care only. Conventional therapy for fit older patients is similar to that for younger patients: the 7 and 3 regimen with standard-dose cytarabine and idarubicin (12 mg/m²), or daunorubicin (60 mg/m²). For patients aged >65 years, high-dose daunorubicin (90 mg/m²) has increased toxicity and is not recommended. A novel liposomal preparation of cytarabine and daunorubicin in a fixed molar ratio may instead be administered to fit patients with AML with myelodysplasia-related changes or arising from MDS. Older patients and those unable to receive intensive therapy due to medical comorbidity may receive repetitive cycles of lower intensity therapy with a hypomethylating agent (HMA; decitabine or azacitidine) or low-dose cytarabine, in combination with daily venetoclax. As noted, targeted IDH1- or IDH2-directed therapy is another consideration. All patients should be considered for clinical trials. Investigational therapy remains the best option for many older patients but especially those with adverse-risk features. (Table 104-6).

With the 7 and 3 regimen, 60–80% of younger and 33–60% of older patients (among those who are candidates for intensive

therapy) with primary AML achieve CR. Response rates around 60% have been similarly reported with the combination of HMA plus venetoclax in older or infirm patient groups. Of patients who do not achieve CR, most have drug-resistant leukemia. Induction death is more frequent with advancing age and medical comorbidity. Patients with refractory disease after induction should be considered for salvage treatments, preferably on clinical trials. Planning for the possibility of allogeneic hematopoietic stem cell transplantation (HCT) for all eligible patients under age 75 years is part of optimal initial AML care. Typically, allogeneic HCT is performed for patients in CR but at risk for relapse, but fit younger patients with primary refractory disease (not in remission after initial induction) have ~15–20% cure rates with allogeneic HCT (after myeloablative conditioning). For this reason, early planning for possible future allogeneic HCT (including human leukocyte antigen [HLA] typing, donor search, etc.) should be part of the initial approach for most AML patients.

POSTREMISSION THERAPY

Induction of a durable first CR (CR1) is critical to long-term survival in AML. However, without further therapy, virtually all CR patients will eventually relapse. Thus, postremission therapy is designed to eradicate residual (typically undetectable) leukemic cells to prevent relapse and prolong survival. As with induction, the type of postremission therapy in AML is selected for each individual patient based on age, fitness, and cytogenetic/molecular risk.

TABLE 104-6 Novel Therapies in Clinical Development in Acute Myeloid Leukemia (AML)

Protein kinase inhibitors	<ul style="list-style-type: none"> FLT3 inhibitors (midostaurin, quizartinib, gilteritinib, crenolanib, sorafenib) KIT inhibitors PI3K/AKT/mTOR inhibitors Aurora and polo-like kinase inhibitors, CDK4/6 inhibitors, CHK1, WEE1, and MPS1 inhibitors SRC and HCK inhibitors Syk inhibitors
Epigenetic modulators	<ul style="list-style-type: none"> New DNA methyltransferase inhibitors (SGI-110) Histone deacetylase (HDAC) inhibitors IDH1 and IDH2 inhibitors DOT1L inhibitors BET-bromodomain inhibitors
Chemotherapeutic agents	<ul style="list-style-type: none"> CPX-351 (liposomal cytarabine and daunorubicin, especially in secondary AML) Vosaroxin Nucleoside analogues
Mitochondrial inhibitors	<ul style="list-style-type: none"> Bcl-2, Bcl-xL, and Mcl-1 inhibitors Caseinolytic protease inhibitors
Therapies targeting oncogenic proteins	<ul style="list-style-type: none"> Fusion transcript targeting EVI1 targeting NPM1 targeting Hedgehog inhibitors (glasdegib)
Antibodies and immunotherapies	<ul style="list-style-type: none"> Monoclonal antibodies against CD33, CD44, CD47, CD123, CLEC12A Immunoconjugates (e.g., gemtuzumab ozogamicin, SGN33A) Bispecific T-cell engagers (BiTEs) and dual affinity retargeting molecules (DARTs) Chimeric antigen receptor (CAR) T cells or genetically engineered T-cell receptor (TCR) T cells Immune checkpoint inhibitors (PD-1/PD-L1, CTLA-4) Vaccines (e.g., WT1)
Therapies targeting AML environment	<ul style="list-style-type: none"> CXCR4 and CXCL12 antagonists Antiangiogenic therapies

Source: Republished with permission of American Society of Hematology, from Diagnosis and management of acute myeloid leukemia in adults: 2017 recommendations from an international expert panel, Döhner H et al. 129:424, 2017; permission conveyed through Copyright Clearance Center, Inc.

The choice between consolidation with chemotherapy or with transplantation is complex and based on age, risk, and practical considerations. In younger patients receiving chemotherapy, postremission therapy with intermediate- or high-dose cytarabine for two to four cycles is standard practice. Higher doses of cytarabine during postremission therapy appear more effective than standard doses (such as are used in induction) for those who do not have adverse-risk genetics. Recent studies suggest that the long-standing practice of high-dose cytarabine (3 g/m², every 12 h on days 1, 3, and 5) may not improve survival over intermediate-dose cytarabine (IDAC; 1–1.5 g/m²) for such patients. Thus, the ELN has recommended IDAC at 1–1.5 g/m², every 12 h, on days 1–3, as the optimal postremission chemotherapy approach for favorable- and intermediate-risk younger patients, for two to four cycles. While high-dose cytarabine may not be necessary, it is important to note that younger, favorable-risk patients have worse outcomes when doses <1 g/m² are used. In contrast to favorable-risk patients, intermediate- or adverse-risk patients should proceed with allogeneic HCT in CR1 when feasible (see transplant discussion below). Because older patients have increased toxicities with higher doses of cytarabine, ELN recommends relatively attenuated cytarabine doses (0.5–1 g/m², every 12 h, on days 1–3) in favorable-risk older patients. There is no

clear value for intensive postremission therapy in non-favorable-risk older patients; allogeneic HCT in CR1 (up to age 75 years) or investigational postremission therapy is recommended. Indeed, postremission therapy is an appropriate setting for introduction of new agents in both older and younger patients (Table 104-6).

For patients treated initially with lower intensity regimens that include venetoclax, the current practice is to continue repetitive cycles of the same combination of agents after remission until disease progression. Therapy often must be abbreviated over time due to cumulative myelotoxicity.

Allogeneic HCT is the best relapse-prevention strategy currently available for AML. Allogeneic HCT is probably best understood as an opportunity for immunotherapy; residual leukemia cells potentially elicit an immunologic response from donor immune cells, the so-called graft-versus-leukemia (GVL) effect. The benefit of GVL in relapse risk reduction, unfortunately, is offset somewhat by increased morbidity and mortality from complications of allogeneic HCT including graft-versus-host disease (GVHD). Given that relapsed AML is typically resistant to chemotherapy, allogeneic HCT in CR1 (e.g., before relapse ever occurs) is a favored strategy. We have often explained to patients that transplant can effectively “eliminate the needle in a haystack, but not a stack of needles.” Transplant is recommended for patients age <75 years who do not have favorable-risk disease and who have an HLA-matched donor (related or unrelated). We also recommend allogeneic HCT in CR1 for patients with intermediate-risk disease (Table 104-3). However, considerable debate exists regarding whether allogeneic HCT in CR1 is a requirement for younger patients with intermediate-risk AML, as one large series from the Medical Research Council reported that such patients have similar outcomes if transplanted only after relapse (and achievement of CR2), sparing some the long-term morbidity of transplantation. That said, allogeneic HCT is generally recommended as soon as possible after CR1 is achieved unless the patient is in a favorable-risk group. Increasingly, patients without HLA-matched donors are considered for alternative donor transplants (e.g., HLA-mismatched unrelated, haploidentical related, and umbilical cord blood) even in CR1. More effective and safe methods of *in vivo* T-cell depletion (i.e., posttransplant cyclophosphamide following mismatched transplantation) have broadened the availability of potential allogeneic HCT donors. Now, virtually any patient with a healthy parent or child (i.e., haploidentical) has an available donor suitable for allogeneic HCT if desired. Long-term outcomes with conventional chemotherapy for older patients are dismal; transplantation for such patients is expanding. Even for older patients, nonrandomized data demonstrate curative potential for older patients in CR1 treated with reduced-intensity conditioning regimens and allogeneic HCT.

Trials comparing allogeneic HCT with intensive chemotherapy or autologous HCT have shown improved duration of remission with allogeneic HCT. However, the relapse risk reduction observed with allogeneic HCT is partially offset by the increase in fatal treatment-related toxicity (GVHD, organ toxicity). Despite this, there is no debate that patients with adverse-risk AML have improved long-term survival with early allogeneic HCT. Alternatively, high-dose chemotherapy with autologous HCT rescue is another postremission approach in non-adverse-risk subsets. Autologous HCT patients receive their own stem cells (collected during remission and cryopreserved), following administration of myeloablative chemotherapy. The toxicity is relatively low with autologous HCT (5% mortality rate), but the relapse rate is higher than with allogeneic HCT due to the absence of the GVL effect. Favorable- and intermediate-risk patients may benefit from autologous HCT more so than adverse-risk patients. Practically speaking, however, autologous HCT in AML patients is less frequently employed currently due to enhanced relapse risk reduction seen with allogeneic HCT and the growing availability of HLA-mismatched donors (in novel transplantation approaches).

Prognostic factors help to select the appropriate postremission therapy in patients in CR1. Our approach includes allogeneic HCT

in CR1 for patients without favorable cytogenetics or genotype. Patients with adverse-risk disease should proceed to allogeneic HCT at CR1 if possible. The decision for allogeneic HCT for younger intermediate-risk patients is complex and individualized as described above; we recommend it when an HLA-matched donor is available. Subsets of patients may benefit from targeted therapy given during remission; emerging data demonstrate survival benefit from incorporation of the FLT3 inhibitor midostaurin, for example, into induction and postremission therapies for patients with *FLT3*-mutated AML. Allogeneic transplantation in CR1 is still recommended for these patients.

For patients in morphologic CR, measurement of MRD remains a very important and challenging research area. Cytogenetics are a mainstay of disease assessment, and persistence of abnormal karyotype (in spite of morphologic CR) is clearly associated with poor clinical outcomes. Immunophenotyping to detect minute populations of blasts or sensitive molecular assays (e.g., reverse transcriptase polymerase chain reaction [RT-PCR]) to detect AML-associated molecular abnormalities (e.g., *NPM1*, *RUNX1/RUNX1T1* and *CBFβ/MYH11* transcripts, *PML/RARA*) can be performed to assess whether MRD is present at sequential time points during or after treatment. Whether emerging next-generation sequencing or serial quantitative assessment using flow or RT-PCR, performed during remission, can effectively direct successful subsequent therapy and improve clinical outcome remains to be determined. Currently, no consensus exists for the optimal MRD measurement technique or its application, although it is increasingly employed in clinical practice. Data suggest that MRD measurement can in some settings be a reliable discriminator between patients who will continue in CR or relapse, but whether subsequent therapy (i.e., allogeneic HCT or additional therapy) can effectively eradicate disease in such patients is not yet clear. However, in the subset of patients with APL, serial RT-PCR (for the *PML/RARA* transcript) is a very useful and reliable tool to detect early relapse and to direct initiation of reinduction therapy prior to onset of overt relapse. Critical in the general understanding of MRD in all disease subsets is the recognition that even patients with undetectable levels of MRD remain at risk for leukemic relapse.

SUPPORTIVE CARE

Measures geared to supporting patients through several weeks of neutropenia and thrombocytopenia are critical to successful AML therapy. Patients with AML should be treated in centers expert in providing supportive care. Multi-lumen central venous catheters should be inserted as soon as newly diagnosed AML patients have been stabilized. They should be used thereafter for administration of intravenous medications/chemotherapy and transfusions, as well as for blood drawing instead of venipuncture during prolonged periods of myelosuppression.

Adequate and prompt blood bank support is critical to therapy of AML. Platelet transfusions should be given as needed to maintain a platelet count $\geq 10,000/\mu\text{L}$. The platelet count should be kept at higher levels in febrile patients and during episodes of active bleeding or DIC. Patients with poor posttransfusion platelet count increments may benefit from administration of ABO-matched platelets or platelets from HLA-matched donors. RBC transfusions should be considered to keep the hemoglobin level $> 70-80 \text{ g/L}$ ($7-8 \text{ g/dL}$) in the absence of active bleeding, DIC, or congestive heart failure, which require higher hemoglobin levels. Blood products leukodepleted by filtration should be used to avert or delay alloimmunization as well as febrile reactions. Blood products may also be irradiated to prevent transfusion-associated GVHD. Cytomegalovirus (CMV)-negative blood products should be used for CMV-seronegative patients who are potential candidates for allogeneic HCT; fortunately, white blood cell filtration is quite effective at reducing CMV exposure as well.

Neutropenia (neutrophils $< 500/\mu\text{L}$ or $< 1000/\mu\text{L}$ and predicted to decline to $< 500/\mu\text{L}$ over the next 48 h) can be part of the initial presentation and/or a side effect of the chemotherapy treatment in

AML patients. Thus, infectious complications remain the major cause of morbidity and death during induction and postremission chemotherapy for AML. Antibacterial (i.e., quinolones) and antifungal (i.e., posaconazole) prophylaxis, especially in conjunction with regimens that cause mucositis, is beneficial. For patients who are herpes simplex virus or varicella-zoster seropositive, antiviral prophylaxis should be initiated (e.g., acyclovir, valacyclovir).

Fever develops in most patients with AML, but infections are documented in only half of febrile patients. Early initiation of empirical broad-spectrum antibacterial and antifungal antibiotics has significantly reduced the number of patients dying of infectious complications (Chap. 74). An antibiotic regimen adequate to treat gram-negative organisms should be instituted at the onset of fever in a neutropenic patient after clinical evaluation, including a detailed physical examination with inspection of the indwelling catheter exit site and a perirectal examination (for perirectal abscess), as well as procurement of cultures and radiographs aimed at documenting the source of fever. Specific antibiotic regimens should be based on institutional antibiotic sensitivity data obtained from where the patient is being treated. Acceptable regimens for empiric antibiotic therapy include monotherapy with imipenem-cilastatin, meropenem, piperacillin/tazobactam, or an extended-spectrum antipseudomonal cephalosporin (cefepime or ceftazidime). The combination of an aminoglycoside with an antipseudomonal penicillin (e.g., piperacillin) or an aminoglycoside in combination with an extended-spectrum antipseudomonal cephalosporin should be considered in complicated or resistant cases. Aminoglycosides should be avoided, if possible, in patients with renal insufficiency. Empirical vancomycin should be added in neutropenic patients with catheter-related infections, blood cultures positive for gram-positive bacteria before final identification and susceptibility testing, hypotension or shock, or known colonization with penicillin/cephalosporin-resistant pneumococci or methicillin-resistant *Staphylococcus aureus*. In special situations where decreased susceptibility to vancomycin, vancomycin-resistant organisms, or vancomycin toxicity is documented, other options including linezolid and daptomycin need to be considered.

Caspofungin (or a similar echinocandin), voriconazole, isavuconazonium, or liposomal amphotericin B should be considered for antifungal treatment if fever persists for 4–7 days following initiation of empiric antibiotic therapy. Although liposomal formulations of amphotericin B have improved the toxicity profile of this agent, use has been limited to situations with high risk of or documented mold infections, especially in those in whom an azole fails. Caspofungin has been approved for empiric antifungal treatment. Voriconazole has also been shown to be equivalent in efficacy and less toxic than amphotericin B; isavuconazonium may also be effective with fewer drug-drug interactions. Antibacterial and antifungal antibiotics should be continued until patients are no longer neutropenic, regardless of whether a specific source has been found for the fever. Unfortunately, this practice likely contributes to development of resistance and increased incidence of nosocomial infections such as *Clostridium difficile* colitis, so great care should be taken preferably in hospital-wide antibiotic surveillance and isolation strategies to reduce these complications. Recombinant hematopoietic growth factors have a limited role in AML; myeloid growth factors may be useful in the postremission setting but are not recommended in induction or for “palliative” care for patients not in remission.

TREATMENT FOR REFRACTORY OR RELAPSED AML

In patients who relapse after achieving CR, the length of first CR is predictive of response to salvage chemotherapy treatment; patients with longer first CR (> 12 months) generally relapse with drug-sensitive disease and have a higher chance of attaining a CR, even with the same chemotherapeutic agents used for first remission induction. Patients with short prior CR duration are at high risk for treatment failure. Similar to patients with refractory disease, patients with relapsed disease are rarely cured by salvage chemotherapy treatments alone. Therefore, patients who eventually

achieve a second CR and are eligible for allogeneic HCT should be transplanted. For patients who relapse after allogeneic HCT, no consensus for best therapy exists; outcomes in this setting are very poor.

Because achievement of a second CR with routine salvage therapies is relatively uncommon, especially in patients who relapse rapidly after achievement of first CR (<12 months), these patients and those lacking HLA-compatible donors or who are not candidates for allogeneic HCT should be considered for innovative approaches on clinical trials. Many new agents are in current testing (Table 104-6). The discovery of novel gene mutations and mechanisms of leukemogenesis that might represent actionable therapeutic targets has prompted the development of many new targeting agents. In addition to kinase inhibitors for *FLT3*-mutated AML, other compounds targeting the aberrant activity of mutant proteins (e.g., IDH1/2 inhibitors) and numerous other biologic mechanisms are being tested in clinical trials. Inhibitors of *FLT3* (gilteritinib), IDH1 (ivosidenib), or IDH2 (enasidenib) are monotherapies for relapsed AML patients who have targetable mutations. Furthermore, approaches with antibodies targeting markers commonly expressed on leukemia blasts (e.g., CD33) or leukemia-initiating cells (e.g., CD123) are also under investigation. Once these compounds have demonstrated safety and activity as single agents, investigation of combinations with other molecular targeting compounds and/or chemotherapy should be pursued.

TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA

APL is a highly curable AML subtype, and ~85% of these patients achieve long-term survival with current approaches. APL has long been shown to be responsive to cytarabine and daunorubicin, but in the past, patients who were treated with these drugs alone frequently died from DIC induced by the release of granule components by the chemotherapy-treated leukemia cells. However, the prognosis of APL patients has changed dramatically with the introduction of tretinoin (*all-trans*-retinoic acid [ATRA]), an oral drug that induces the differentiation of leukemic cells bearing the t(15;17), where disruption of the *RARA* gene encoding a retinoid acid receptor occurs. ATRA decreases the frequency of DIC but often produces another complication called the APL (differentiation) syndrome. Occurring within the first 3 weeks of treatment, it is characterized by fever, fluid retention, dyspnea, chest pain, pulmonary infiltrates, pleural and pericardial effusions, and hypoxemia. The syndrome is related to adhesion of differentiated neoplastic cells to the pulmonary vasculature endothelium. Glucocorticoids, chemotherapy for cytoreduction, and/or supportive measures can be effective for management of the APL syndrome. Temporary discontinuation of ATRA is necessary in cases of severe APL syndrome (i.e., patients developing renal failure or requiring admission to the intensive care unit due to respiratory distress). The mortality rate of this syndrome is ~10%. APL syndrome may also occur, less commonly, with arsenic trioxide (ATO) in APL.

In adults with low-risk APL (low leukocyte count at presentation), ATRA (45 mg/m²/d) plus ATO (0.15 mg/kg/d) was recently compared to ATRA plus concurrent idarubicin chemotherapy. ATRA/ATO was superior and is the new standard of care for such patients. CR rates in low-risk disease approach 100%, with excellent long-term survival. Notably, patients with high-risk APL (defined as leukocyte count >10,000/ μ L) must be uniquely treated, as they require immediate cytoreduction with chemotherapy due to life-threatening APL syndrome and rapidly rising leukocyte count after initiation of ATRA. High-risk patients are at increased risk for induction death due to this syndrome as well as increased frequency of hemorrhagic complications (related to DIC).

Assessment of residual disease by RT-PCR amplification of the t(15;17) chimeric gene product *PML-RARA* following the final cycle of treatment is important. Disappearance of the signal is associated with long-term disease-free survival; its persistence or reemergence invariably predicts relapse. Sequential monitoring of RT-PCR for *PML-RARA* is now considered standard for postremission monitoring of APL, at least in high-risk patients.

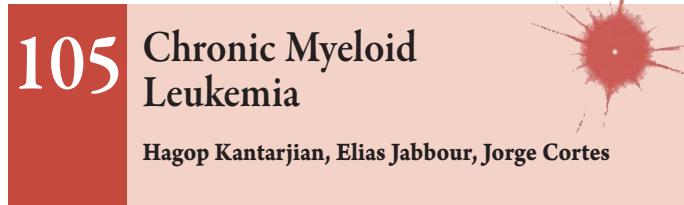
Patients in molecular, cytogenetic, or clinical relapse should be salvaged with ATO with or without ATRA; in patients who were treated with ATRA plus chemotherapy in the front-line setting, ATO-based therapy at relapse produces meaningful responses in up to 85% of patients. Although experience with relapsed APL in patients who received ATO during initial induction is limited (given that few relapses occur in low-risk patients and widespread use of ATO during first-line therapy is relatively new), ATO remains the preferred reinduction therapy for patients who relapse, although the duration of prior remission should be a factor in this choice. Achievement of CR2 should be followed by consolidation with autologous HCT (for patients who achieve RT-PCR-negative status). In the minority who do not achieve negative RT-PCR or who relapse again, allogeneic HCT may still be potentially curative.

ACKNOWLEDGEMENT

Clara Bloomfield, an important contributor to the field and to this chapter in past editions, passed away since the publication of the 20th edition. Material from prior versions of this chapter on which she was an author have been retained here.

FURTHER READING

- ARBER DA et al: Acute myeloid leukaemia (AML) with recurrent genetic abnormalities, in *World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues*, 4th revised ed. Geneva, Switzerland: World Health Organization; 2016.
- DINARDO CD et al: Venetoclax combined with decitabine or azacitidine in treatment-naïve, elderly patients with acute myeloid leukemia. *Blood* 133:7, 2019.
- DÖHNER H et al: Diagnosis and management of acute myeloid leukemia in adults: 2017 recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 129:424, 2017.
- JAISWAL S, EBERT BL: Clonal hematopoiesis in human aging and disease. *Science* 366:eaan4673, 2019.
- JONGEN-LAVRENIC M et al: Molecular minimal residual disease in acute myeloid leukemia. *N Engl J Med* 378:1189, 2018.
- LO-COCO F et al: Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N Engl J Med* 369:111, 2013.
- PAPAEMMANILOU E et al: Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med* 374:2209, 2016.
- PERL AE et al: Gilteritinib or chemotherapy for relapsed or refractory *FLT3*-mutated AML. *N Engl J Med* 381:1728, 2019.
- POLLYEA DA et al: Enasidenib, an inhibitor of mutant IDH2 proteins, induces durable remissions in older patients with newly diagnosed acute myeloid leukemia. *Leukemia* 33:2575, 2019.
- ROBOZ GJ et al: Ivosidenib induces deep durable remissions in patients with newly diagnosed IDH1-mutant acute myeloid leukemia. *Blood* 135:463, 2020.
- STONE RM et al: Midostaurin plus chemotherapy for acute myeloid leukemia with a *FLT3* mutation. *N Engl J Med* 377:454, 2017.



Chronic myeloid leukemia (CML) is a clonal hematopoietic myeloproliferative stem cell neoplasm. The disease is driven by the *BCR/ABL1* chimeric gene that codes for a constitutively active tyrosine kinase, resulting from a reciprocal balanced translocation between the long arms of chromosomes 9 and 22, t(9;22)(q34.1;q11.2), known as the

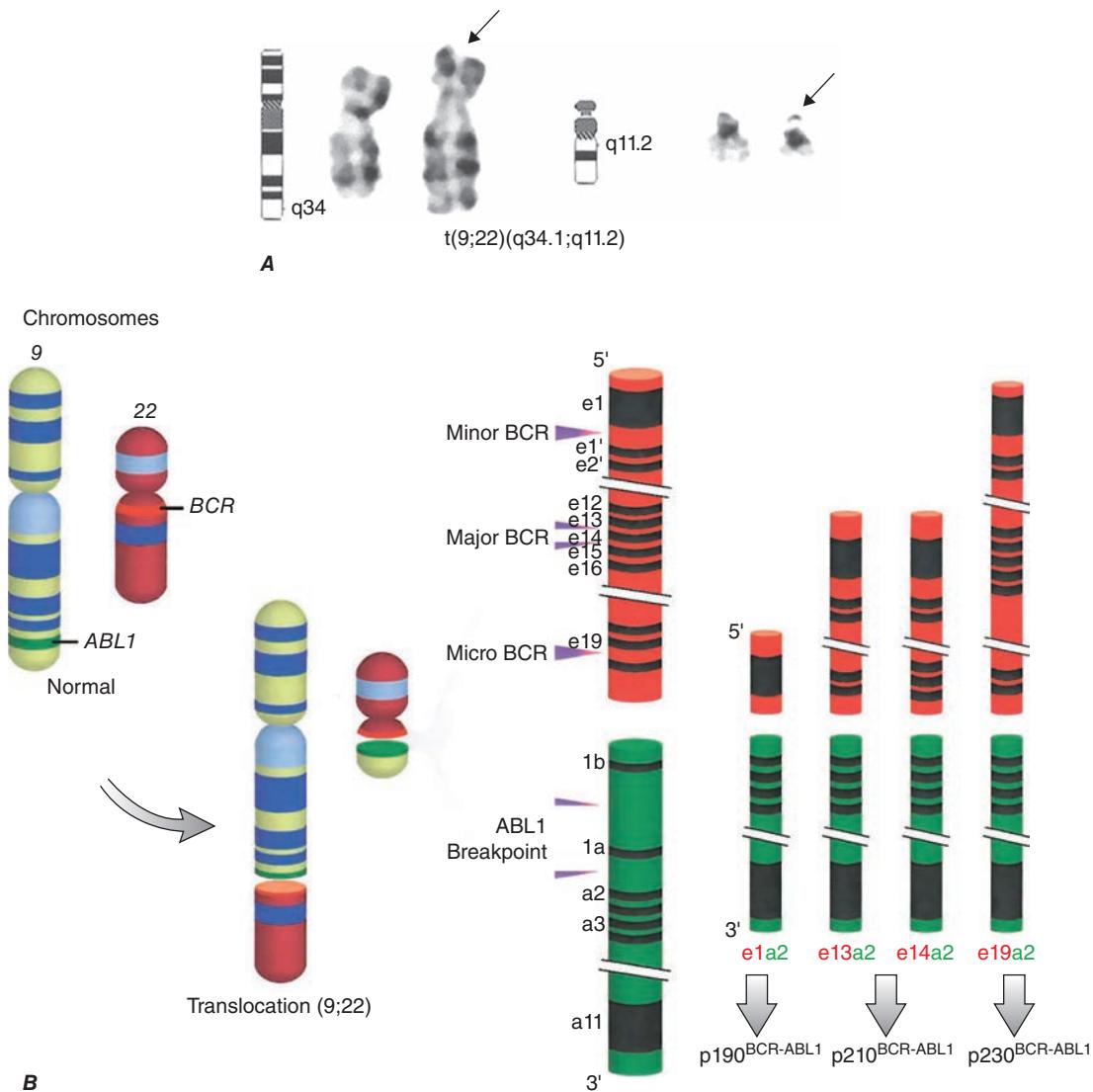


FIGURE 105-1 **A.** The Philadelphia (Ph) chromosome cytogenetic abnormality. **B.** Breakpoints in the long arms of chromosome 9 (ABL1 locus) and chromosome 22 (BCR regions) result in at least three different BCR-ABL1 oncprotein messages, p210^{BCR-ABL1} (most common message in chronic myeloid leukemia [CML]), p190^{BCR-ABL1} (present in two-thirds of patients with Ph-positive acute lymphoblastic leukemia; rare in CML), and p230^{BCR-ABL1} (rare in CML and associated with an indolent course). Other rearrangements (e.g., e14a3, e14a3) are less common. (© 2013 The University of Texas MD Anderson Cancer Center.)

Philadelphia chromosome (Ph) (Fig. 105-1). Untreated, the course of CML is typically biphasic or triphasic, with an early indolent or chronic phase, followed often by an accelerated phase and a terminal blastic phase. Before the era of BCR-ABL1 tyrosine kinase inhibitors (TKIs), the median survival in CML was 3–7 years, and the 10-year survival rate was 30% or less. Introduced into standard CML therapy in 2000, TKIs have revolutionized the treatment, natural history, and prognosis of CML. Today, the estimated 10-year survival rate with imatinib mesylate, the first BCR-ABL1 TKI approved, is greater than 85% and approaches that of the general population. Allogeneic stem cell transplantation (SCT), a curative approach but one that involves more risks, is now offered as second- or third-line therapy after failure of TKIs.

INCIDENCE AND EPIDEMIOLOGY

CML accounts for ~15% of all cases of leukemia. There is a slight male predominance (male-to-female ratio 1.6:1). The median age at diagnosis is 55–65 years. It is uncommon in children; only 3% of patients with CML are younger than 20 years, although in recent years, a higher proportion of young patients are diagnosed. The incidence of CML increases gradually with age, with a steeper increase after the age of 40–50 years. The annual incidence of CML is 1.6 cases per 100,000 individuals. In the United States, this translates into about 8500–9000 new cases per year. The incidence of CML has not changed

over several decades. By extrapolation, the worldwide annual incidence of CML is about 200,000 cases. With a median survival of 3–6 years before 2000, the disease prevalence in the United States was ~30,000 cases. With TKI therapy, the annual mortality has been reduced from 10–20% to about 2%. Therefore, the prevalence of CML is expected to continue to increase. Based on an estimated annual mortality of 2% and an incidence of 8500 cases per year, the plateau prevalence of CML is estimated to be reached at ~425,000 in the United States ($8500 \times 100/2$) by about 2040, with full TKI optimal treatment penetration. The worldwide prevalence will depend on the treatment penetration of TKIs and their effect on reduction of worldwide annual mortality. Ideally, with full TKI treatment penetration, the worldwide prevalence should plateau at 35 times the incidence, or ~9–10 million patients. These estimates are all based on extrapolations from the incidence and prevalence of CML in the United States, as well as an estimated annual mortality of 2% with modern TKI therapy; they could vary considerably if the estimates were to change.

ETIOLOGY

There are no familial associations in CML. The risk of developing CML is not increased in monozygotic twins or in relatives of patients with CML. No etiologic agents are incriminated, and no associations exist with exposures to benzene or other toxins, fertilizers, insecticides, or

viruses. CML is not a frequent secondary leukemia following therapy of other cancers with alkylating agents and/or radiation. Exposure to ionizing radiation (e.g., nuclear accidents, radiation treatment for ankylosing spondylitis or cervical cancer) has increased the risk of CML, which peaks at 5–10 years after exposure and is dose-related. The median time to development of CML among atomic bomb survivors was 6.3 years. Following the Chernobyl accident, no increase in the incidence of CML was reported, suggesting that larger dose exposures of radiation are required to cause CML. Because of adequate protection, the risk of CML has not increased among individuals working in the nuclear industry or among radiologists.

■ PATHOPHYSIOLOGY

The *t(9;22)(q34.1;q11.2)* is present in >90% of classical CML cases. It results from a balanced reciprocal translocation between the long arms of chromosomes 9 and 22. It is present in hematopoietic cells (myeloid, erythroid, megakaryocytes, and monocytes; less often mature B lymphocytes; rarely mature T lymphocytes, but not stromal cells), but not in other cells in the human body. As a result of the genetic translocation, DNA sequences from the cellular oncogene *ABL1* are juxtaposed to the major breakpoint cluster region (*BCR*) gene on chromosome 22, generating a hybrid oncogene, *BCR/ABL1*. Depending on the breakpoint site in the major *BCR* region on chromosome 22 (e13 or e14), two main messenger RNA transcripts occur, e13a2 (previously b2a2) and e14a2 (previously b3a2). Both of them encode for a novel oncoprotein of molecular weight 210 kDa, referred to as p210^{BCR-ABL1} (Fig. 105-1B). This oncoprotein exhibits constitutive kinase activity that leads to excessive proliferation and reduced apoptosis of CML cells, endowing them with a growth advantage over their normal counterparts. Over time, normal hematopoiesis is suppressed, but normal stem cells can persist and reemerge following effective therapy, for example with TKIs. In most instances of Ph-positive acute lymphoblastic leukemia (ALL) and in rare cases of CML, the breakpoint in *BCR* is more centromeric, in a region called the minor *BCR* region (*mBCR*). As a result, a shorter sequence of *BCR* is fused to *ABL1*, with a consequent e1a2 transcript and a smaller BCR-ABL1 oncoprotein, p190^{BCR-ABL1}. When occurring in Ph-positive CML, this translocation is associated with a worse outcome. A rarer breakpoint in *BCR* occurs telomeric to the major *BCR* region in the *micro-BCR* (μ -*BCR*) region. It juxtaposes a larger fragment of the *BCR* gene to *ABL1* and produces an e19a2 transcript and a larger p230^{BCR-ABL1} oncoprotein (associated with a more indolent CML course). Other rearrangements (based on different breakpoints in the *ABL* region), such as e13a3 or e14a3 (also resulting in a p210^{BCR-ABL1} oncoprotein), occur much less frequently. These are not readily identifiable nor quantifiable with the routine polymerase chain reaction (PCR) probes, thus producing falsely negative PCR levels on follow-up studies if not tested at diagnosis.

The constitutive activation of *BCR/ABL1* results in autophosphorylation and activation of multiple downstream pathways that affect gene transcription, apoptosis, stromal adherence, skeletal organization, and degradation of inhibitory proteins. These transduction pathways involve RAS, mitogen-activated protein (MAP) kinases, signal transducers and activators of transcription (STAT), phosphatidylinositol-3-kinase (PI3k), MYC, and others. These interactions are mostly mediated through tyrosine phosphorylation and require binding of BCR-ABL1 to adapter proteins such as GRB-2, CRK, CRK-like (CRK-L) protein, and Src homology containing proteins (SHC). Most BCR-ABL1 TKIs bind to the BCR-ABL1 ATP-binding domain, inhibiting its kinase activity, preventing the activation of transformation pathways, and inhibiting downstream signaling. As a result, proliferation of CML cells is inhibited and apoptosis induced, allowing the reemergence of normal hematopoiesis. An additional layer of complexity is related to differences in signal transduction between CML-differentiated cells and early progenitors. Beta-catenin, Wnt1, Foxo3a, transforming growth factor β , interleukin-6, PP2A, SIRT1, and others have been implicated in CML stem cell survival. *ABL1* also has a myristoyl site that functions as a negative regulator of its kinase activity. This site and its negative regulatory activity are lost upon fusion with *BCR*. Novel *ABL1* inhibitors (e.g., asciminib) bind this myristoyl site

and restore the lost inhibitory activity. Mutations in other cancer-associated genes may also occur at diagnosis, most frequently in *ASXL1*, *IKZF1*, and *RUNX1*; their presence is associated with worse response to therapy and a higher risk of transformation to blastic phase.

Experimental models have established the causal relationship between the *BCR/ABL1* rearrangement and the development of CML. In animal models, expression of *BCR/ABL1* in normal hematopoietic cells produced CML-like disorders or lymphoid leukemia, demonstrating the leukemogenic potential of *BCR/ABL1* as a single oncogenic abnormality. Other models, however, suggest the need for a "second hit."

The cause of the *BCR/ABL1* rearrangement is unknown. Molecular techniques that detect *BCR/ABL1* at a level of 1 in 10⁸ cells identify this molecular abnormality in the blood of up to 25% of normal adults and 5% of infants, but 0% of cord blood samples. This suggests that *BCR/ABL1* is not sufficient to cause overt CML in the overwhelming majority of individuals in whom it occurs. Because CML develops in only 1.6 of 100,000 individuals annually, additional molecular events or poor immune recognition of the rearranged cells may contribute to overt CML.

CML is defined by the presence of the *BCR/ABL1* fusion gene in a patient with a myeloproliferative neoplasm. In some patients with a typical morphologic picture of CML, the Ph chromosome is not detectable by standard G-banding karyotype, but fluorescence in situ hybridization (FISH) and/or molecular studies (PCR) detect *BCR/ABL1*. These patients have a course similar to patients with Ph-positive CML and respond to TKI therapy. Many of the remaining patients have atypical morphologic or clinical features and have other diseases, such as atypical CML, chronic myelomonocytic leukemia, and myelodysplastic/myeloproliferative neoplasms (MDS/MPN). These individuals do not respond to TKI therapy and usually have a poor prognosis with a median survival of about 2–3 years. Detection of mutations in the granulocyte colony-stimulating factor receptor (*CSF3R*) in chronic neutrophilic leukemia (80% of cases) and in some cases of atypical CML (5–10% of cases), mutations in *SETBP1* in atypical CML (25% of cases), and mutations in *SF3B1* in MDS/MPN with ringed sideroblasts and marked thrombocytosis (MDS/MPN-RS-T; 50–70% of cases, associated with longer median survival of 7 years vs 3.3 years with wild-type *SF3B1*) supports the notion that these are distinct molecular and biologic entities. Patients with chronic neutrophilic leukemia or atypical CML whose disease is associated with *CSF3R* mutation may respond well to ruxolitinib (a JAK2 inhibitor) therapy (complete response in 50–60% of such patients).

The events associated with the transition of CML from a chronic to accelerated-blastic phase are poorly understood. Characteristic chromosomal abnormalities such as a double Ph, trisomy 8, isochromosome 17 or deletion of 17p (loss of *TP53*), 20q-, translocations involving 3q26, and others may be seen with disease acceleration. Molecular events associated with transformation include mutations in *TP53*, retinoblastoma 1 (*RBL*), myeloid transcription factors like *RUNX1*, and cell cycle regulators like *p16*. A plethora of other mutations or functional abnormalities have been implicated in blastic transformation, but no unifying theme has emerged other than the fact that *BCR/ABL1* itself induces genetic instability that favors the acquisition of additional molecular defects and eventually results in blastic transformation. One critical effect of TKIs is to stabilize the CML genome, leading to a reduced transformation rate. In particular, the previously observed sudden blastic transformations (i.e., abrupt transformation to blastic phase in a patient who had been in cytogenetic response) have become uncommon, occurring rarely in younger patients in the first 1–2 years of TKI therapy (usually sudden lymphoid blastic transformations). Sudden transformations beyond the third year of TKI therapy are rare in patients who continue on TKI therapy. Moreover, the course of CML is now frequently more indolent in patients treated with TKI, even without cytogenetic response, compared to previous experience with hydroxyurea/busulfan, suggesting a definite clinical benefit of continued inhibition of the kinase activity.

Among patients developing resistance to TKIs, several resistance mechanisms have been observed. The most clinically relevant one is the development of *ABL1* kinase domain mutations that may prevent the

binding of TKIs to the catalytic site (ATP-binding site) of the kinase or maintain the kinase activity despite the presence of a TKI. More than 100 *ABL1* kinase domain mutations have now been described, many of which confer relative or absolute resistance to imatinib. Consequently, second-generation (i.e., dasatinib, nilotinib, bosutinib) and third-generation (ponatinib) TKIs were developed, the latter with significant efficacy against T315I, a “gatekeeper” mutation that prevents binding of and causes resistance to all other currently available TKIs. Asciminib, olveremabtinib (HQPI351) and other novel TKIs under development are also active against the T315I mutation.

■ CLINICAL PRESENTATION

The presenting signs and symptoms in CML depend on the availability of and access to health care, including physical examinations and screening tests. In the United States, because of the wider access to health care screening and physical examinations, 50–60% of patients are diagnosed on routine blood tests and have minimal symptoms at presentation, such as fatigue. In geographic locations where access to health care is more limited, patients often present with high CML burden including splenomegaly, anemia, and related symptoms (abdominal pain, weight loss, fatigue), associated with a higher frequency of high-risk CML. Presenting findings in patients diagnosed in the United States are shown in **Table 105-1**.

Symptoms Most patients with CML (90%) present in the indolent or chronic phase. Depending on the timing of diagnosis, patients are often asymptomatic (if the diagnosis is discovered during health care screening tests). Common symptoms, when present, are manifestations of anemia and splenomegaly. These may include fatigue, malaise, weight loss (if high leukemia burden), or early satiety and left upper quadrant pain or masses (from splenomegaly). Less common presenting findings include thrombotic or hyperviscosity-related events from severe leukocytosis or thrombocytosis. These include priapism, cardiovascular complications, myocardial infarction, venous thrombosis, visual disturbances, dyspnea and pulmonary insufficiency, drowsiness, loss of coordination, confusion, or cerebrovascular accidents.

TABLE 105-1 Presenting Signs and Symptoms of Newly Diagnosed Philadelphia Chromosome-Positive Chronic Myeloid Leukemia in Chronic Phase

PARAMETER	PERCENTAGE
Age ≥60 years (median)	40–50 (55–65)
Female gender	35–45
Splenomegaly	30
Hepatomegaly	5–10
Lymphadenopathy	5
Other extramedullary disease	2
Hemoglobin <10 g/dL	10–15
Platelets	
>450 × 10 ⁹ cells/L	30–35
<100 × 10 ⁹ cells/L	3–5
White blood cells ≥50 × 10 ⁹ cells/L	35–40
Marrow	
≥5% blasts	5
≥5% basophils	10–15
Peripheral blood	
≥3% blasts	8–10
≥7% basophils	10
Cytogenetic clonal evolution other than the Philadelphia chromosome	4–5
Sokal risk	
Low	60–65
Intermediate	25–30
High	10

Manifestations of bleeding diatheses include retinal hemorrhages, gastrointestinal bleeding, and others. Patients who present with, or progress to, the accelerated or blastic phases frequently have additional symptoms including unexplained fever, significant weight loss, severe fatigue, bone and joint pain, bleeding and thrombotic events, and infections.

Physical Findings Splenomegaly is the most common physical finding, occurring in 20–70% of patients depending on health care screening frequency. Other less common findings include hepatomegaly (5–10%), lymphadenopathy (5–10%), and extramedullary disease (skin or subcutaneous lesions). The latter indicates CML transformation if a biopsy confirms predominance of blasts. Other physical findings are manifestations of complications of high tumor burden described earlier (e.g., cardiovascular, cerebrovascular, bleeding). High basophil counts may be associated with histamine overproduction causing pruritus, diarrhea, flushing, and even gastrointestinal ulcers.

Hematologic and Marrow Findings In untreated CML, leukocytosis ranging from 10–500 × 10⁹/L is common. The peripheral blood differential shows left-shifted hematopoiesis with predominance of neutrophils and the presence of bands, myelocytes, metamyelocytes, promyelocytes, and blasts (usually ≤5%). Basophils and/or eosinophils are frequently increased. Thrombocytosis is common, but thrombocytopenia is rare and, when present, suggests a worse prognosis, disease acceleration, or an unrelated etiology. Anemia is present in one-third of patients. Cyclic oscillations of counts are noted in 10–20% of patients without treatment. Biochemical abnormalities include a low leukocyte alkaline phosphatase score and high levels of vitamin B₁₂, uric acid, lactic dehydrogenase, and lysozyme. The presence of unexplained and sustained leukocytosis, with or without splenomegaly, should lead to a marrow examination and cytogenetic analysis.

The bone marrow is hypercellular with marked myeloid hyperplasia and a high myeloid-to-erythroid ratio of 15–20:1. Marrow blasts are typically 5% or less; when higher, they carry a worse prognosis or represent transformation to accelerated phase (if they are ≥15%). Increased reticulin fibrosis (detected with silver stain) is common, with 30–40% of patients demonstrating grade 3–4 reticulin fibrosis. This was considered adverse in the pre-TKI era. With TKI therapy, reticulin fibrosis resolves in most patients and is not an indicator of poor prognosis. Collagen fibrosis (Wright-Giemsa stain) is rare at diagnosis. Disease progression with a “spent phase” of myelofibrosis (myelophthisis, or burnt-out marrow) was a relatively common end-stage CML condition with busulfan therapy (20–30%); it is extremely rare now with TKI therapy.

Cytogenetic and Molecular Findings The diagnosis of CML is straightforward and depends on documenting the t(9;22) (q34.1;q11.2), which is identified by G-banding in 90% of cases. This is known as the Philadelphia chromosome (initially identified in Philadelphia as a minute chromosome, later identified to be chromosome 22) (Fig. 105-1). Some patients (~10%) may have complex translocations (complex variant Ph) involving three or more chromosomes including chromosomes 9 and 22 and one or more additional chromosomes. Others may have a “masked Ph,” involving translocations between chromosome 9 and a chromosome other than 22 (but molecularly showing the *BCR/ABL1* rearrangement; known as simple variant Ph). The prognosis of these patients and their response to TKI therapy are similar to those in patients with Ph. About 5–10% of patients may have additional chromosomal abnormalities (ACAs) in the Ph-positive cells at diagnosis. These usually involve trisomy 8, a double Ph, isochromosome 17 or 17p deletion, 20q-, or others. This is referred to as cytogenetic clonal evolution and was historically a sign of adverse prognosis, particularly when trisomy 8, double Ph, or chromosome 17 abnormalities were noted. A less common abnormality involving chromosome 3q26.2 occurs with disease progression and carries a poor prognosis.

Techniques such as FISH and PCR are now used to aid in the diagnosis of CML. They are more sensitive to estimate the CML burden in patients on TKI therapy. They can be done on peripheral blood and thus are more convenient to patients. Patients with CML at diagnosis

should have a FISH analysis to quantify the percentage of Ph-positive cells, if FISH is used to replace marrow cytogenetic analysis in monitoring response to therapy. FISH will not detect additional chromosomal abnormalities (clonal evolution); thus, a cytogenetic analysis is recommended at the time of diagnosis. In addition, 10–15% of patients may develop chromosomal abnormalities in Ph-negative metaphases after responding to TKIs. These abnormalities may carry a worse prognosis but are not detected by FISH unless already identified and FISH is used to follow them. Molecular studies at diagnosis are important to document the type and presence of *BCR-ABL1* transcripts to avoid spurious “undetectable” *BCR-ABL1* transcripts on follow-up studies, with the false impression of a complete molecular response. The presence of the Philadelphia chromosome with “negative” PCR with standard methodology should prompt investigation of atypical transcripts.

Both FISH and PCR studies can be falsely positive at low levels or falsely negative because of technical issues. Therefore, a diagnosis of CML must always rely on a marrow analysis with routine cytogenetics. The diagnostic bone marrow confirms the presence of the Ph chromosome, detects clonal evolution, and quantifies the percentage of marrow blasts and basophils. In 10% of patients, the percentage of marrow blasts and basophils can be significantly higher than in the peripheral blood, conferring poorer prognosis or even representing disease transformation.

Monitoring patients on TKI therapy by cytogenetics, FISH, and PCR has become an important standard practice to assess response to therapy, emphasize compliance, evaluate possible treatment resistance, identify the need to change TKI therapy, and determine the need to assess for kinase domain mutations. Because of the decreasing reliance of bone marrow aspirations to monitor response, equivalence has been established to correlate cytogenetic results with PCR values. These are not absolute correlations but provide adequate guidance. A partial cytogenetic response is defined as the presence of 35% or less Ph-positive metaphases by routine cytogenetic analysis. This is roughly equivalent to *BCR-ABL1* transcripts by the International Scale (IS) of 10% or less. A complete cytogenetic response refers to the absence of Ph-positive metaphases (0% Ph positivity). This is approximately equivalent to *BCR-ABL1* transcripts (IS) of 1% or less. A major molecular response (MMR or MR3) refers to *BCR-ABL1* transcripts (IS) ≤0.1%, or roughly a 3-log or greater reduction of *BCR-ABL1* transcripts from a standardized baseline. MR4 refers to *BCR-ABL1* transcripts (IS) ≤0.01%, and MR4.5 (deep molecular response) refers to *BCR-ABL1* transcripts (IS) ≤0.0032%, roughly equivalent to a 4.5-log reduction or greater of transcripts.

Findings in CML Transformation Progression of CML is usually associated with leukocytosis resistant to therapy, increasing anemia, fever and constitutional symptoms, and increased blasts and basophils in the peripheral blood or marrow. Criteria of accelerated-phase CML, historically associated with median survival of <1.5 years, include the presence of 15% or more peripheral blasts, 30% or more peripheral blasts plus promyelocytes, 20% or more peripheral basophils, cytogenetic clonal evolution (presence of chromosomal abnormalities in addition to Ph), and thrombocytopenia <100 × 10⁹/L (unrelated to therapy). About 5–10% of patients present with de novo accelerated phase or blastic phase. The prognosis of de novo accelerated phase with TKI therapy has improved significantly, with an estimated 8-year survival rate of 75%. The median survival of accelerated phase evolving from chronic phase has also improved from a historical median survival of 18 months to an estimated 4-year survival rate of 70% on TKI therapy. Therefore, the criteria for accelerated-phase CML should be revisited because most clinical criteria defining accelerated phase have lost much of their prognostic significance. Blastic-phase CML is defined by the presence of 30% or more peripheral or marrow blasts or the presence of sheets of blasts in extramedullary disease (usually skin, soft tissues, or lytic bone lesions). Blastic-phase CML is commonly myeloid (60%) but can present uncommonly as erythroid, promyelocytic, monocytic, or megakaryocytic. Lymphoid blastic phase occurs in about 25% of patients. Lymphoblasts are terminal deoxynucleotide transferase positive and peroxidase negative (although occasionally

with low positivity up to 3–5%) and express lymphoid markers (CD10, CD19, CD20, CD22). However, they also often express myeloid markers (50–80%), resulting in diagnostic challenges. Proper immunophenotypic diagnosis is important because lymphoid blastic-phase CML is quite responsive to anti-ALL-type chemotherapy (e.g., hyper-CVAD [cyclophosphamide, vincristine, doxorubicin, and dexamethasone]) in combination with TKIs (complete response rate 70%; median survival 3 years; high rates of bridging to allogeneic SCT and possible cure).

■ PROGNOSIS AND CML COURSE

Before the imatinib era, the annual mortality in CML was 10% in the first 2 years and 15–20% thereafter. The median survival in CML was 3–7 years (with hydroxyurea-busulfan and interferon α). Without a curative option of allogeneic SCT, the course of CML was toward transformation to, and death from, accelerated or blastic phases for most patients as the rate of complete cytogenetic response with interferon was low. Even apparent disease stability was unpredictable, with some patients demonstrating sudden transformation to a blastic phase. With imatinib therapy, the annual mortality in CML has decreased to 1–2% in the first 20 years of observation. More than half of the deaths are from conditions other than CML, such as old age, comorbidities, accidents, suicides, other cancers, and other medical conditions (e.g., infections, surgical procedures). The estimated 10-year survival rate is 86%, or 92% if only CML-related deaths are considered (Fig. 105-2). The course of CML has also become quite predictable. In the first 2 years of TKI therapy, rare sudden transformations are still reported (1–2%), usually lymphoid blastic transformations that respond to combinations of chemotherapy and TKIs followed by allogeneic SCT. These may be explained by the intrinsic mechanisms of sudden transformation already existing in the CML clones before the start of therapy that were not amenable to TKI inhibition, in particular imatinib. Second-generation TKIs (nilotinib, dasatinib, bosutinib) used as frontline therapy have reduced the incidence of transformation in the first 2–3 years from 6–8% with imatinib to 2–5% with second-generation TKIs. Disease transformation to accelerated or blastic phase is rare with continued TKI therapy, estimated at <1% annually in years 4–10 of follow-up on the original imatinib trials. Patients usually develop resistance in the form of cytogenetic resistance or relapse, followed by hematologic relapse and subsequent transformation, rather than the previously feared sudden transformations without the warning signals of cytogenetic-hematologic relapse.

Before the imatinib era, several pretreatment prognostic factors predicted for worse outcome in CML and have been incorporated into prognostic models and staging systems. These have included older age, significant splenomegaly, anemia, thrombocytopenia or thrombocytosis, high percentages of blasts and basophils (and/or eosinophils), marrow fibrosis, interstitial deletions in the long arm of chromosome 9, clonal evolution, and others. Different risk models and staging systems, derived from multivariate analyses, were proposed to define different risk groups. As with the introduction of cisplatin into testicular cancer therapy, the introduction of TKIs into CML therapy has decreased or, in some instances, eliminated the prognostic impact of most of these prognostic factors and the significance of the CML models (e.g., Sokal, Hasford, European Treatment and Outcome Study [EUTOS]). Treatment-related prognostic factors have emerged as the most important prognostic factors in the era of imatinib therapy. Achievement of complete cytogenetic response has become the major therapeutic endpoint and is the only endpoint associated with improvement in survival. Achievement of MMR or MR3 is associated with decreased risk of events (relapse) and CML transformation but has not been associated with survival prolongation among patients with complete cytogenetic response. This may be due to the survival benefit conferred by the achievement of complete cytogenetic response, which approximates normal life expectancy, and to the efficacy of salvage TKI therapies, which are and should be implemented at the first evidence of cytogenetic relapse. Achievement of undetectable *BCR-ABL1* transcripts (complete molecular response [CMR]) or deep molecular response (DMR; defined as MR4 or MR4.5), particularly when sustained (>2–5 years), may offer the possibility of treatment-free remission and may

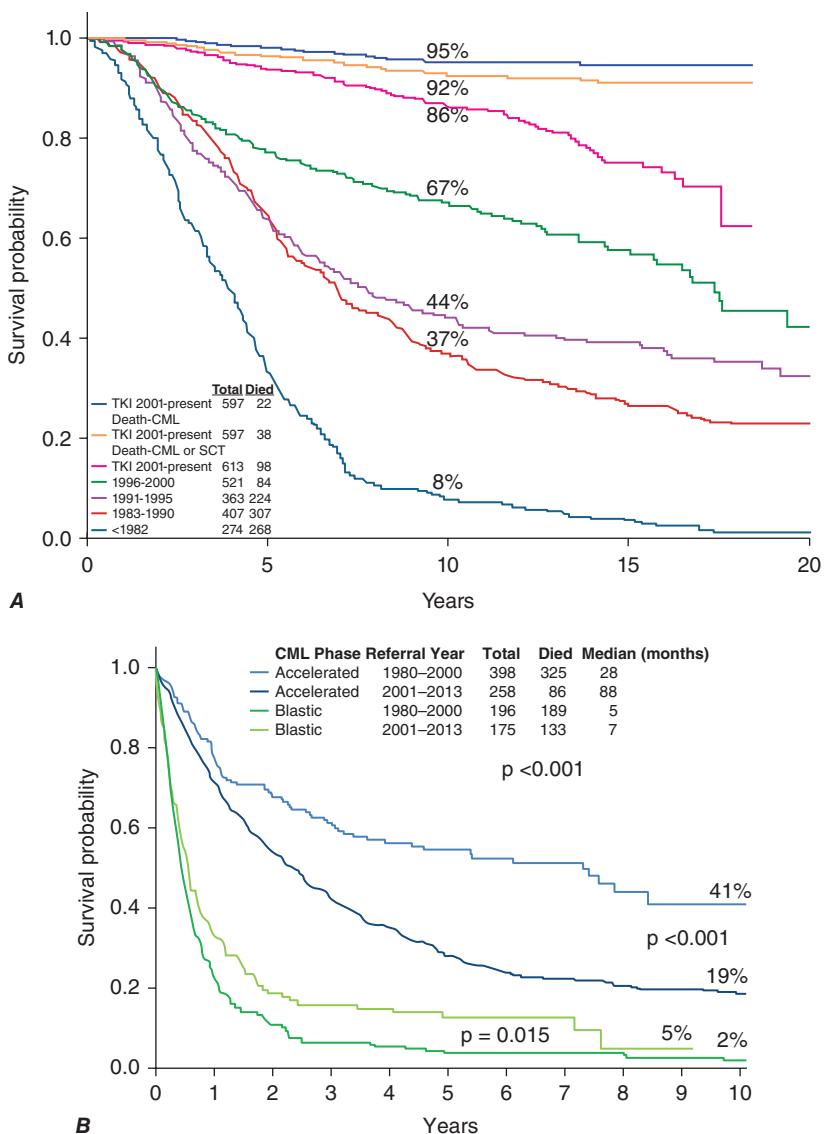


FIGURE 105-2 **A.** Survival in newly diagnosed chronic-phase chronic myeloid leukemia (CML) by era of therapy (MD Anderson Cancer Center experience from 1965 to present). Top blue curve is survival with tyrosine kinase inhibitors (TKIs), accounting for only CML-related deaths. The orange curve (second from top) accounts for deaths related to CML or CML treatment complications (e.g., deaths following allogeneic stem cell transplant [SCT]). The red curve (third from top) is survival including all deaths regardless of causality (old age, car accidents, suicide, gun shots, second cancers, complications of unrelated surgeries, infections, others). The difference in the denominators, 613 minus 597 cases, is because 16 deaths were from unknown/undocumented causes (outside MD Anderson and no good tracking for cause of death). **B.** Survival in patients with accelerated- and blastic-phase CML referred to MD Anderson Cancer Center by era of therapy, demonstrating the significant survival benefit in the TKI era in accelerated-phase CML but the modest benefit in blastic-phase CML. Referred cases included de novo and post-chronic-phase transformations.

allow temporary therapy interruption in women pursuing pregnancy. The lack of achievement of MMR or DMR should not be considered as “failure” of a particular TKI therapy and/or an indication to change the TKI or to consider allogeneic SCT.

Long-term updates of randomized trials suggest that second-generation TKIs and imatinib are similarly effective in lower-risk CML; second-generation TKIs may offer a therapeutic advantage among patients with high-risk CML.

TREATMENT

Chronic Myeloid Leukemia

Since 2001, six drugs have been approved by the U.S. Food and Drug Administration (FDA) for the treatment of CML. These include five oral TKIs: imatinib (Gleevec, Glivec), nilotinib (Tasigna), dasatinib (Sprycel), bosutinib (Bosulif), and ponatinib (Iclusig). Dasatinib, nilotinib, and bosutinib are referred to as second-generation TKIs; ponatinib is referred to as a third-generation TKI. Nilotinib is

similar in structure to imatinib but 30 times more potent. Dasatinib and bosutinib inhibit the SRC family of kinases in addition to ABL1, with dasatinib reported to be 300 times more potent and bosutinib 30–50 times more potent than imatinib. In contrast to all other TKIs, bosutinib has no activity against c-Kit or platelet-derived growth factor receptor (PDGFR). Ponatinib is highly effective against wild-type and mutant BCR/ABL1 clones. It is also the only available BCR-ABL1 TKI active against T315I, a gatekeeper mutation resistant to the other four ATP-competitive TKIs (Table 105-2). Ponatinib also inhibits vascular endothelial growth factor receptor (VEGFR), which may be at least partly responsible for the high incidence of hypertension observed with this agent (Table 105-2). Imatinib 400 mg orally daily, nilotinib 300 mg orally twice a day (on an empty stomach), dasatinib 100 mg orally daily, and bosutinib 400 mg orally daily are approved for frontline therapy of CML. Dasatinib 50 mg orally daily is as effective in frontline therapy as 100 mg daily, and significantly less toxic. All four are also approved for salvage therapy (nilotinib 400 mg twice daily; bosutinib 500 mg daily; others at the same dose as frontline therapy), in addition to ponatinib

TABLE 105-2 Medical Therapeutic Options in Chronic Myeloid Leukemia

AGENT (BRAND NAME)	APPROVED INDICATIONS	DOSE SCHEDULE	NOTABLE TOXICITIES
Imatinib mesylate (Gleevec)	All phases	400 mg daily	See text
Dasatinib (Sprycel)	All phases	First-line: 100 mg daily Salvage: 100 mg daily in chronic phase; 140 mg daily in transformation	Myelosuppression; pleural and pericardial effusions; pulmonary hypertension
Nilotinib (Tasigna)	All phases except blastic phase	First-line: 300 mg twice daily Salvage: 400 mg twice daily	Diabetes; arterio-occlusive disease; pancreatitis
Bosutinib (Bosulif)	All phases	First line: 400 mg daily Salvage: 500 mg daily	Diarrhea; liver toxicity; renal dysfunction
Ponatinib (Iclusig)	Optimal TKI if T315I mutation Failure of ≥2 tyrosine kinase inhibitors	45 mg daily (may consider lower starting doses in the future, e.g., 30 mg daily). (Lower the dose to 15 mg daily once a complete cytogenetic response is achieved).	Skin rashes (10–20%); pancreatitis (5%); arterio-occlusive disease (10–20%); systemic hypertension (10–15%)
Omacetaxine mepesuccinate (Synribo)	Failure ≥2 tyrosine kinase inhibitors	1.25 mg/m ² subcutaneously twice daily for 14 days of induction; 7 days of maintenance every month (may consider shorter dose schedules, 7 days of induction, 2–5 days of maintenance)	Myelosuppression

(45 mg daily). Ponatinib 45 mg daily may be associated with serious side effects: arterio-occlusive events, pancreatitis, hypertension, and skin rashes. A response-directed dose adjusted regimen, with a starting dose of 45 mg and reduction to 15 mg once a cytogenetic response is achieved, has resulted in a reduced incidence of arterio-occlusive events and has become standard. Imatinib, dasatinib (140 mg daily), bosutinib, and ponatinib are also approved for the treatment of CML in transformation (accelerated and blastic phase), whereas nilotinib is only approved for chronic and accelerated phase. The sixth approved drug is omacetaxine (Synribo), a protein synthesis inhibitor with presumed more selective inhibition of the synthesis of the BCR-ABL1 oncoprotein. It is approved for the treatment of chronic- and accelerated-phase CML after failure of two or more TKIs, at 1.25 mg/m² subcutaneously twice a day for 14 days for induction and for 7 days for consolidation-maintenance. The main adverse event of omacetaxine is prolonged myelosuppression; thus, many experts use shorter schedules (e.g., omacetaxine 5–7 days induction and 2–5 days maintenance), often combined with a TKI (Table 105-2).

Imatinib, dasatinib, bosutinib, and nilotinib are all acceptable frontline therapies in CML. The long-term results of imatinib are very favorable. The 10-year follow-up results show a cumulative complete cytogenetic response rate (occurring at least once) of 83%, with 60–65% of patients being in complete cytogenetic response at 5-year follow-up. The estimated 10-year survival rate is ~85%. Among patients continuing on imatinib, the annual rate of transformation to accelerated-blastic phase in years 4–8 is <1%. In three randomized studies, one comparing nilotinib 300 mg twice daily or 400 mg twice daily with imatinib (ENESTnd), another comparing dasatinib 100 mg daily with imatinib (DASISION), and a third comparing bosutinib 400 mg daily with imatinib (BFORE), the second-generation TKIs were associated with better outcomes in early surrogate endpoints, including higher rates of complete cytogenetic responses (85–87% vs 77–82%), MMRs (5-year rates 76–77% vs 60–64%), and MR4.5 (5-year rates 42–53% vs 31–33%), with lower rates of transformation to accelerated and blastic phase (2–5% vs 7%). However, no study has shown a survival benefit with second-generation TKIs. This may be because the rate of complete cytogenetic response is ultimately similarly high with imatinib versus second-generation TKIs, and also because sequential therapy with TKIs (following close observation and treatment change at progression) provides highly effective therapy for most patients; this ensures adequate long-term outcome despite relapse or intolerance after initial therapy.

Salvage therapy in chronic phase with dasatinib, nilotinib, bosutinib, or ponatinib is associated with complete cytogenetic response rates of 30–60%, depending on the salvage status (cytogenetic vs

hematologic relapse), prior response to other TKIs, number of prior TKIs used, and the mutations at the time of relapse. Complete cytogenetic responses are generally durable, particularly in the absence of clonal evolution. Ponatinib is the only TKI active in the setting of T315I mutation, with complete cytogenetic response rates of 50–70% among patients who have received two or more TKIs. The estimated 5-year survival rates with new TKIs as salvage are 70–75% (compared with <50% before their availability). For example, with dasatinib salvage after imatinib failure in chronic-phase CML, the estimated 7-year rate of major molecular was 46%, the estimated 7-year survival rate was 65%, and progression-free survival rate was 42%. Thus, TKIs in the salvage setting have already reduced the annual mortality from the historical rate of 10–15% to ≤5%.

The goal of CML therapy is survival prolongation. The achievement of treatment-free remission (TFR) status has become a therapeutic goal of increased interest (sustained DMR or CMR after discontinuation of TKI therapy). In current practice, with the availability of appropriate TKI therapy and with compliance, monitoring, and changing of TKI therapy as indicated by response/resistance and side effects, patients can have a near-normal life expectancy, with a “relative” survival similar to that of the general population. Therefore, in standard practice, achievement and maintenance of a complete cytogenetic response are the aims of therapy, because complete cytogenetic response is the only outcome associated with survival prolongation. Lack of achievement of an MMR (protects against events; associated with longer event-free survival) or of DMR (offers the potential of treatment discontinuation and of TFR) should not be considered indications to change TKI therapy or to consider allogeneic SCT. A general practice rule is to continue the particular TKI chosen at the most tolerable dose schedule not associated with grade 3–4 side effects or with bothersome chronic side effects, for as long as possible, until either cytogenetic relapse or the persistence of unacceptable side effects. These two factors (i.e., cytogenetic relapse and intolerable side effects) are the indicators of “failure” of a particular TKI therapy. A second emerging general practice rule is that patients with CML should always receive daily TKI therapy throughout their lifetime (chronic, transformation), either alone (chronic) or in combinations (possibly for those in transformation, although combinations not formally approved), except perhaps in situations of “molecular cure” (TFR; elective discontinuation of TKI if DMR sustained for >2 to 5 years, followed by close monitoring) or after allogeneic SCT with undetectable disease.

Because of the increasing prevalence of CML (cost of TKI therapy) and the emerging evidence of possible organ toxicities with long-term use (e.g., renal with imatinib and bosutinib; arterio-occlusive with nilotinib, dasatinib, and ponatinib), a goal of therapy

of increasing interest in CML is to achieve eradication of the disease (molecular “cure” or TFR) that is prolonged and durable, with recovery of nonneoplastic, nonclonal hematopoiesis off TKI therapy. The first step toward this aim is to obtain the highest rates of DMR lasting for at least 2 or more years. This is currently achievable in about 25–30% of patients treated with imatinib and in 40–45% of patients treated with second-generation TKIs. Approximately 50–60% of those who meet these criteria and discontinue therapy remain free from therapy and in DMR-MMR. As a result, TFR rates are estimated to be about 15–20% after imatinib therapy and 25–30% after second-generation TKIs.

Recommendations provided by the National Comprehensive Cancer Network (NCCN) and by the European LeukemiaNet (ELN) propose optimal/expected, suboptimal/warning, and failure response scenarios at different time points of TKI treatment duration. Unfortunately, they may have been misinterpreted in current practice because oncologists often report that their aim of treatment is the achievement of MMR and disease eradication. Significantly, a substantial proportion of oncologists consider a change of TKI therapy in a patient in complete cytogenetic response if they note loss of MMR (increase of *BCR-ABL1* transcripts [IS] from $\leq 0.1\%$ to $> 0.1\%$). This perception may be the result of confusion regarding the aims of the NCCN and ELN guidelines, which have been updated often as a result of maturing data and have multiple treatment endpoint considerations. Although such endpoints may have been suggested as possible criteria for failure or suboptimal response, it is important to emphasize that no randomized study has yet shown that a change of TKI treatment in patients with complete cytogenetic response because of a loss of MMR, versus changing at the time of cytogenetic relapse, improves survival or other long-term outcomes. This is likely because of the high efficacy of salvage TKI therapy at the time of cytogenetic relapse.

Side effects of TKIs are generally mild to moderate, although with long-term TKI therapy, they could affect the patient’s quality of life. Serious side effects occur in <5–10% of patients. With imatinib therapy, common mild to moderate side effects include fluid retention, weight gain, nausea, diarrhea, skin rashes, periorbital edema, bone or muscle aches, fatigue, and others (rates of 10–20%). In general, second-generation TKIs are associated with lower rates of these bothersome adverse events. However, dasatinib 100 mg daily is associated with higher rates of myelosuppression (20–30%), particularly thrombocytopenia, with pleural (10–25%) or pericardial effusions ($\leq 5\%$), and with pulmonary hypertension (<5%). A lower dose of dasatinib (50 mg daily instead of 100 mg daily) used in frontline CML therapy has resulted in similar efficacy and a lower incidence of serious side effects (pleural effusions <5%, myelosuppression <10%). Nilotinib is associated with higher rates of hyperglycemia (10–20%), pruritus and skin rashes, hyperbilirubinemia (typically among patients with Gilbert’s syndrome and mostly of no clinical consequences), and headaches. Nilotinib is also associated with occasional instances of pancreatitis (<5%). Nilotinib 300–400 mg twice daily is associated with a 10-year cumulative incidence of cardiovascular complications of 15–25%. Bosutinib is associated with higher rates of liver toxicity, renal dysfunction, and early and self-limited gastrointestinal adverse events, particularly diarrhea (70–85%). Occasionally, the gastrointestinal symptoms mimic chronic severe enterocolitis, which reverses with treatment discontinuation. Ponatinib 45 mg daily is associated with higher rates of serious skin rashes (10–15%), pancreatitis (10%), elevations of amylase/lipase (10%), and systemic hypertension (50–60%; severe in 20%). Arterio-occlusive events (cardiovascular, cerebrovascular, and peripheral arterial) have been reported with most TKIs. The incidence appears to be highest with ponatinib, but both nilotinib and dasatinib are associated with these events at an incidence significantly higher than imatinib. Among the TKIs, bosutinib is associated with the lowest incidence of cardiovascular events. Nilotinib and dasatinib may cause prolongation of the QTc interval; therefore, they should be evaluated cautiously in patients with prolonged QTc

interval on electrocardiogram (>470–480 ms), and drugs given for other medical conditions should have relatively smaller or no effects on QTc. These side effects can often be dose-dependent and are generally reversible with treatment interruptions and dose reductions. Dose reductions can be individualized. However, the lowest estimated effective doses of TKIs (from different studies and treatment practices) are imatinib 100–200 mg daily; nilotinib 150 mg twice daily or 200 mg daily; dasatinib 20 mg daily; bosutinib 200–300 mg daily; and ponatinib 15 mg daily.

With long-term follow-up, rare but clinically relevant serious toxicities are emerging. Renal dysfunction and occasionally renal failure (creatinine elevations $> 2\text{--}3 \text{ mg/dL}$) are observed in 2–3% of patients, more frequently with imatinib and bosutinib than other TKIs, and usually reverse with TKI discontinuation and/or dose reduction. Rarely, patients may develop TKI-related peripheral neuropathy or even central neurotoxicities that are misdiagnosed as dementia or Alzheimer’s disease; they may reverse slowly after TKI discontinuation. Pulmonary hypertension has been reported with dasatinib (<1–2%) and should be considered in a patient with shortness of breath and a normal chest x-ray (echocardiogram with emphasis on measurement of pulmonary artery pressure). This may be reversible with dasatinib discontinuation and occasionally the use of sildenafil citrate. Systemic hypertension has been observed more often with ponatinib. Hyperglycemia and occasionally diabetes have been noted more frequently with nilotinib. Finally, mid- and small-vessel arterio-occlusive and vasospastic events have been reported at low but significant rates with nilotinib and ponatinib and should be considered possibly TKI-related and represent indications to interrupt or reduce the dose of the TKI. These events include angina, coronary artery disease, myocardial infarction, peripheral arterial occlusive disease, transient ischemic attacks, cerebral vascular accidents, Raynaud’s phenomenon, and accelerated atherosclerosis. Although these events are uncommon (<5%) (10-year cumulative rates of 15% with nilotinib 300 mg BID and 20–25% with 400 mg BID, compared with <5% with imatinib), they are clinically significant for the patient’s long-term prognosis and occur at significantly higher rates than in the general population, particularly among patients with other risk factors for such events. Serious arterio-occlusive and vasospastic events are more common with ponatinib 45 mg daily (5-year rates 20%).

Discontinuation of TKIs and Treatment-Free Remissions Several studies have confirmed that TKI discontinuation among patients who achieve DMR (MR4.5) for longer than 2–3 years can result in TFR rates of 40–60%. Discontinuation of TKI therapy after 5+ years of CMR is associated with TFR rates of 70–80% or greater. Since the incidence of durable MR4.5 (*BCR-ABL* transcripts [IS] $\leq 0.0032\%$) is 30–60%, ~15–30% of all patients with CML on TKI therapy may achieve TFR. This approach is ready for community practice provided it is done under optimal conditions. These include the following: patients must have low or intermediate Sokal risk CML in first chronic phase (no evidence or history of transformation), with history of quantifiable *BCR-ABL1* transcripts (e13a2, e14a2), on long-term TKI therapy (5–8+ years), with documented DMR for >2–3 years (assessed every 6 months during this time span and with a PCR with adequate sensitivity), and should be monitored at referral centers that offer rigorous testing of residual CML disease. Patients must also be compliant to frequent monitoring (PCR studies every 1–2 months for the first 6 months, then every 2 months until 2 years and every 3–6 months thereafter).

ALLOGENEIC STEM CELL TRANSPLANT

Allogeneic SCT, a curative modality in CML, is associated with long-term survival rates of 40–60% when implemented in chronic phase. It is associated with early (1-year) mortality rates of 5–30%. Although the 5- to 10-year survival rates were reported to be ~50–60% (and considered as cure rates), ~10–15% of patients die in the subsequent 1–2 decades from subtle long-term complications of the transplant (rather than from CML relapse). These are related

to chronic graft-versus-host disease (GVHD), organ dysfunction, development of second cancers, occasional late relapses, and hazard ratios for mortality higher than in the normal population. Other significant morbidities include infertility, chronic immune-mediated complications, cataracts, hip necrosis, and other morbidities affecting quality of life. The cure and early mortality rates in chronic-phase CML are also associated with several factors: patient age, duration of chronic phase, whether the donor is related or unrelated, degree of matching, preparative regimen, and others. In accelerated-phase CML, the cure rates with allogeneic SCT are 30–50%, depending on the definition of accelerated disease. Patients with clonal evolution as the only criterion have cure rates of up to 40–50%. Patients undergoing allogeneic SCT in second chronic phase have cure rates of 40–50%. The cure rates with allogeneic SCT in blastic-phase CML are ≤20%. Post-allogeneic SCT strategies are now implemented in the setting of molecular or cytogenetic relapse or in hematologic relapse/transformation. These include the use of TKIs for prevention or treatment of relapse, donor lymphocyte infusions, and second allogeneic SCTs, among others. TKIs appear to be highly successful at reinducing cytogenetic/molecular remissions in the setting of cytogenetic or molecular relapse after allogeneic SCT.

Choice and Timing of Allogeneic SCT Allogeneic SCT was considered first-line CML therapy before 2000. The maturing positive experience with TKIs has now relegated its use to after first-line TKI failures. An important question is the optimal timing and sequence of TKIs and allogeneic SCT (whether allogeneic SCT should be used as second- or third-line therapy). Among patients who present with or evolve to blastic phase, combinations of chemotherapy and TKIs should be used to induce remission, followed by allogeneic SCT as soon as possible. The same applies to patients who evolve from chronic to accelerated phase. Patients with de novo accelerated-phase CML may do well with long-term TKI therapy (estimated 8-year survival rate 75%); the timing of allogeneic SCT depends on their optimal response to TKI (achievement of complete cytogenetic response). Among patients who relapse in chronic phase, the treatment sequence depends on several factors: (1) patient age and availability of appropriate donors; (2) risk of allogeneic SCT; (3) presence or absence of clonal evolution and mutations; (4) patient's prior history and comorbidities; and (5) patient and physician preferences (**Table 105-3**). Patients with T315I mutations at relapse should be offered ponatinib and considered for allogeneic SCT particularly if in blastic phase and perhaps also in accelerated phase (because of the short follow-up with ponatinib). Patients with mutations involving Y253H, E255K/V, and F359V/C/I respond better to dasatinib or bosutinib. Patients with mutations involving V299L, T315A, and F317L/F/I/C respond better to nilotinib. Comorbidities such as diabetes, hypertension, pulmonary hypertension, chronic lung disease, cardiac conditions, and pancreatitis may influence the choice for or against a particular TKI. Patients with clonal evolution, unfavorable mutations, or lack of major/complete cytogenetic response within 1 year of salvage TKI therapy have short remission durations and should consider allogeneic SCT as more urgent in the setting of salvage. Patients without clonal evolution or mutations at relapse and who achieve a complete cytogenetic response with TKI salvage have long-lasting complete remissions and may delay the option of allogeneic SCT to third-line therapy. Finally, older patients (age 65–70 years or older) and those with high risk of mortality with allogeneic SCT may forgo this curative option for several years of disease control in chronic phase with or without cytogenetic response (Table 105-3). In emerging nations, where generic imatinib is now available at the annual price of \$400–3000, frontline imatinib is a cost-effective therapy. However, second-line therapy with allogeneic SCT, a one-time curative option with a cost of \$20,000–100,000, may be considered (in preference to second-generation TKIs—annual cost above \$40,000–100,000) as a more cost-effective national health

TABLE 105-3 General Suggestions Regarding the Use of Tyrosine Kinase Inhibitors (TKIs) and Allogeneic Stem Cell Transplantation (SCT) in Chronic Myeloid Leukemia (CML)

CML PHASE	USE OF TKI	CONSIDERATION OF ALLOGENEIC SCT
Accelerated or blastic	Interim therapy to achieve minimal CML burden	As soon as possible (exception: de novo accelerated phase)
T315I mutation	Ponatinib to achieve minimal CML burden	Depends on longer term follow-up results of ponatinib efficacy
Imatinib failure in chronic phase; no clonal evolution, no mutations, good initial response; no T315I	Second-line TKIs long term	Third-line after second-line TKI failures
Clonal evolution or mutations, or no cytogenetic response to second-line TKI	Interim therapy with alternative second-generation TKI or ponatinib to achieve minimal CML burden	Second-line
Older patients (≥65–70 years) after imatinib failure in chronic phase	Salvage TKIs as longer-term therapy	May forgo allogeneic SCT in favor of good quality of life and survival in chronic phase
Imatinib failure; emerging nation	—	Second-line: curative, one-time cost \$20,000–100,000 (vs >\$40,000–100,000/year with TKI)

Note: Mutations involving Y253H, E255K/V, or F359V/C/I: prefer dasatinib or bosutinib. Mutations involving V299L, T315A, or F317L/F/I/C: prefer nilotinib.

care strategy in CML. Table 105-3 summarizes a general guidance to the choice of TKIs versus allogeneic SCT.

MONITORING THERAPY IN CML

Achievement of complete cytogenetic response by 12 months of imatinib therapy and its persistence later, the only consistent prognostic factor associated with prolonged survival, is now the main therapeutic endpoint in CML. Failure to achieve a complete cytogenetic response by 12 months or occurrence of later cytogenetic or hematologic relapse is considered as treatment failure and an indication to change therapy. Because salvage therapy with other TKIs may re-establish good outcome, it is important to ensure patient compliance to continued TKI therapy and change therapy when cytogenetic relapse is confirmed unless this is related to non-adherence. Patients on frontline imatinib therapy should be closely monitored until documentation of complete cytogenetic response, at which time they can be monitored every 6 months with peripheral blood PCR, or more frequently (e.g., every 3 months), if there are concerns about changes in *BCR-ABL1* transcripts. Cytogenetic relapse on imatinib is an indication of treatment failure and need to change TKI therapy. Mutational analysis in this instance helps in the selection of the next TKI and identifies mutations in 30–50% of patients. Mutational studies by standard Sanger sequencing (which is the technique currently available in most clinical laboratories) in patients in complete cytogenetic response (in whom there may be concerns of increasing *BCR-ABL1* transcripts) identify mutations in ≤5% and are therefore not indicated. Earlier response has been identified as a prognostic factor for long-term outcome, including achievement of partial cytogenetic response (*BCR-ABL1* transcripts ≤10%) by 3–6 months of therapy. Failure to achieve such a response has been associated with significantly worse survival.

The use of second-generation TKIs (dasatinib, bosutinib, nilotinib) as frontline therapy changed the monitoring approach slightly. Patients are expected to achieve major cytogenetic response (or *BCR-ABL1* transcripts ≤10%) by 3–6 months of therapy. Failure to do so is associated with worse event-free survival, transformation

rates, and survival. However, the 3- to 5-year estimated survival among such patients is still high, ~80–90%, which is better than what would be anticipated if such patients were offered allogeneic SCT at that time. Changes of therapy for patients with “slow” response have not been proven to be of long-term benefit compared to changes when more obvious signs of resistance appear. Thus, slow response to therapy is considered a warning signal, but it is not known whether changing therapy to other TKIs at that time would improve longer-term outcome.

TREATMENT OF ACCELERATED AND BLASTIC PHASES

Patients in accelerated or blastic phase may receive therapy with TKIs, preferably second- or third-generation TKIs (dasatinib, nilotinib, bosutinib, ponatinib), alone or in combination with chemotherapy, to reduce the CML burden, before undergoing allogeneic SCT. Response rates (major hematologic) with single-agent TKIs range from 30 to 50% in accelerated phase and from 20 to 30% in blastic phase. Cytogenetic responses, particularly complete cytogenetic responses, are uncommon (10–30%) and transient in blastic phase. Studies of TKIs in combination with chemotherapy show that combined TKI-chemotherapy strategies increase the response rates and their durability and improve survival. This is particularly true in CML lymphoid blastic phase, where the combination of anti-ALL chemotherapy with TKIs results in complete response rates of 70% and median survival times of 3 years (compared with historical response rates of 40–50% and median survival times of 12–18 months). This allows many patients to undergo allogeneic SCT in a state of minimal CML burden or second chronic phase, which are associated with higher probability of long-term survival. In CML nonlymphoid blastic phase, anti-acute myeloid leukemia chemotherapy combined with TKIs results in CR rates of 30–50% and median survival times of 9–12 months (compared with historical response rates of 20–30% and median survival times of 3–5 months). In accelerated phase, response to single TKIs is significant in conditions where “softer” accelerated phase criteria are considered (e.g., clonal evolution alone, thrombocytosis alone, significant splenomegaly or resistance to hydroxyurea, but without evidence of high blast and basophil percentages). In accelerated phase, combinations frequently include TKIs with low-intensity chemotherapy such as low-dose cytarabine, decitabine, interferon α , hydroxyurea, or others.

OTHER TREATMENTS AND SPECIAL THERAPEUTIC CONSIDERATIONS

Interferon α Interferon α is considered in combination with TKIs (an investigational approach), sometimes after CML failure on TKIs, occasionally in patients during pregnancy, or as part of investigational strategies with TKIs to eradicate residual molecular disease.

Chemotherapeutic Agents Hydroxyurea remains a safe and effective agent (at daily doses of 0.5–10 g) to reduce initial CML burden, as a temporary measure in between definitive therapies, or in combination with TKIs to sustain complete hematologic or cytogenetic responses. Busulfan is often used in allogeneic SCT preparative regimens. Because of its side effects (delayed myelosuppression, Addison-like disease, pulmonary and cardiac fibrosis, myelofibrosis), it is now rarely used in the chronic management of CML. Low-dose cytarabine, decitabine, anthracyclines, 6-mercaptopurine, 6-thioguanine, thiotepa, anagrelide, and other agents are sometimes useful in different CML settings to control the disease burden.

Others Splenectomy is now seldom considered to alleviate symptoms of massive splenomegaly and/or hypersplenism. Splenic irradiation is rarely used, if at all, because of the postirradiation adhesions and complications. Leukapheresis is occasionally used in patients presenting with extreme leukocytosis and leukostatic complications. Single doses of high-dose cytarabine or high doses

of hydroxyurea, with tumor lysis management, may be as effective and less cumbersome.

Special Considerations Women with CML who become pregnant should discontinue TKI therapy immediately. Among 125 babies delivered to women with CML who discontinued imatinib therapy as soon as the pregnancy was known, three babies were born with neurologic, skeletal, and renal malformations, suggesting the teratogenicity of imatinib known from animal studies. A similar experience has been reported with dasatinib, where the incidence of malformations was reported to be higher, 10–12%. There are no or little data with other TKIs. Control of CML during pregnancy can be managed with leukapheresis for severe symptomatic leukocytosis in the first trimester and with hydroxyurea subsequently until delivery. There are reports of successful pregnancies and deliveries of normal babies with interferon α therapy and registry studies in essential thrombocytosis of its safety, but interferon α has side effects that may be troublesome during pregnancy, can be antiangiogenic, and may increase the risk of spontaneous abortions.

Approximately 10–15% of patients on TKI therapy may develop chromosomal abnormalities in the Ph-negative cells. These may involve loss of chromosome Y, trisomy 8, 20q-, chromosome 5 or 7 abnormalities, and others. Most chromosomal abnormalities disappear spontaneously and may be indicative of the genetic instability of the hematopoietic stem cells that predisposes the patient to develop CML in the first place. Rarely (in <1% of instances), abnormalities involving chromosomes 5 or 7 may be truly clonal and evolve into myelodysplastic syndrome or acute myeloid leukemia. This is thought to be part of the natural course of patients in whom CML was suppressed and who live long enough to develop other hematologic malignancies.

GLOBAL ASPECTS OF CML

Routine physical examinations and blood tests in the United States and advanced countries result in early detection of CML in most patients. About 50–70% of patients with CML are diagnosed incidentally, and high-risk CML as defined by prognostic models (e.g., Sokal risk groups) is found in only 10% of patients. This is not the same situation in emerging nations where most patients are diagnosed following evaluation for symptoms and many present with high tumor burden, such as massive splenomegaly, and advanced phases of CML (high-risk CML documented in 20–30%). Therefore, the prognosis of such patients on TKI therapy may be worse than the published experience.

The high cost of TKI therapies (annual costs of \$90,000–140,000 in the United States; lower but variable in the rest of the world) makes the general affordability of such treatments difficult. Although TKI treatment penetration is high in nations where cost of therapy is not an issue (e.g., Sweden, European Union), it may be less so in other nations, even in advanced ones like the United States, where out-of-pocket expenses may be prohibitive to a subset of patients. Although the estimated 10-year survival in CML is >85% in single-institution studies (e.g., MD Anderson Cancer Center), in national studies in countries with TKI affordability (Sweden) (Figs. 105-2 and 105-3) or in clinical trials (where all patients have access to TKIs throughout their care), the estimated 10-year survival worldwide, even 16 years after the introduction of TKI therapies, is likely to be <50%. The Surveillance, Epidemiology, and End Results (SEER) data from the United States report an estimated 5-year survival rate of 60% in the era of TKIs. It appears that the treatment penetration of imatinib and other TKIs into CML therapy worldwide is still not optimal.

The current high cost of TKI therapies poses two additional considerations. The first are the treatment pathways and guidelines in nations where TKIs may not be affordable by patients or the health care system. In these conditions, there are trends of pathways advocating allogeneic SCT as frontline or second-line therapy (i.e., after imatinib failure; as a one-time cost of \$20,000–100,000) despite the associated mortality and morbidities. The second is the choice of frontline TKI therapy. Imatinib is now available in generic forms at affordable costs (\$400–10,000 per

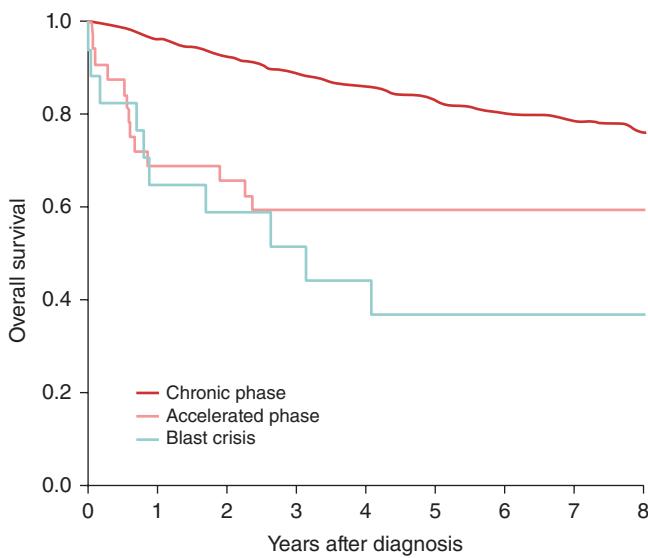


FIGURE 105-3 Survival in chronic (CP), accelerated (AP), and blastic crisis (BC) phases of chronic myeloid leukemia (CML) in the population-based Swedish national registry study. The accelerated- and blastic-phase cases are de novo presentations. The favorable outcome with de novo blastic phase may be due to use of 20% blasts or more to define blastic phase. (With permission from Dr. Martin Hoglund, Swedish CML Registry, 2013.)

year). Dasatinib is available in generic forms in many geographies. Safe and effective generic TKIs may become preferred frontline and salvage therapies in CML, precluding the necessity of an allogeneic SCT in first salvage in poorer nations.

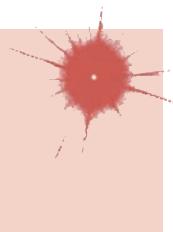
FURTHER READING

- CORTES JE et al: Final 5-year study results of DASISION: The dasatinib versus imatinib study in treatment-naïve chronic myeloid leukemia patients trial. *J Clin Oncol* 34:2333, 2016.
- CORTES JE et al: Bosutinib versus imatinib for newly diagnosed chronic myeloid leukemia: Results from the randomized BFORE Trial. *J Clin Oncol* 36:231, 2018.
- CORTES JE et al: Ponatinib efficacy and safety in Philadelphia chromosome-positive leukemia: Final 5-year results of the phase 2 PACE trial. *Blood* 132:393, 2018.
- HOCHHAUS A et al: Long-term outcomes of imatinib treatment for chronic myeloid leukemia. *N Engl J Med* 376:917, 2017.
- HOCHHAUS A et al: European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia* 34:966, 2020.
- JABBOUR E, KANTARJIAN H: Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. *Am J Hematol* 95:691, 2020.
- KANTARJIAN H et al: Long-term outcomes with frontline nilotinib versus imatinib in newly diagnosed chronic myeloid leukemia in chronic phase: ENESTnd 10-year analysis. *Leukemia* 35:440, 2021.
- NAQVI K et al: Long-term follow-up of lower dose dasatinib (50 mg daily) as frontline therapy in newly diagnosed chronic-phase chronic myeloid leukemia. *Cancer* 126:67, 2020.
- SASAKI K et al: Conditional survival in patients with chronic myeloid leukemia in chronic phase in the era of tyrosine kinase inhibitors. *Cancer* 122:238, 2016.
- SAUSSELE S et al: Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): A prespecified interim analysis of a prospective, multicentre, non-randomised trial. *Lancet Oncol* 19:747, 2018.
- SHIH YT et al: Treatment value of second-generation BCR-ABL1 tyrosine kinase inhibitors compared with imatinib to achieve treatment-free remission in patients with chronic myeloid leukaemia: A modelling study. *Lancet Haematol* 6:e398, 2019.

106

Acute Lymphoid Leukemia

Dieter Hoelzer



In acute lymphoblastic leukemia (ALL), the malignant clone arises from hematopoietic progenitors in the bone marrow or lymphatic system resulting in an increase of immature nonfunctioning leukemic cells. Infiltration of bone marrow leads to anemia, granulocytopenia, and thrombocytopenia with the clinical manifestations of fatigue, weakness, infection, and hemorrhage. These symptoms are more often the reason a patient first seeks medical advice rather than consequences of tumor bulk, such as lymph node enlargement, hepatosplenomegaly caused by leukemic infiltration, or symptoms of the central nervous system (meningeosis leukemica).

INCIDENCE AND AGE

ALL is the most frequent neoplastic disease in children with an early peak at the age of 3–4 years. The incidence in adults ranges from 0.7 to 1.8/100 000 per year, being somewhat higher in adolescents and young adults (AYAs), decreasing in adults, but increasing again in elderly people. Thus, Philadelphia chromosome-positive ALL (Ph+ ALL; *BCR/ABL* translocation) is observed in half of elderly B-lineage patients. The frequency of immunologic, cytogenetic, and genetic subtypes changes substantially with age.

ETIOLOGY

The etiology of acute leukemias is unknown. Internal and external factors influence the incidence of leukemia. Exposure to ionizing radiation or to chemicals, including prior chemotherapy, is associated with an increased risk of developing leukemia, more often observed in acute myeloid leukemia (AML). However, increasingly, secondary ALLs have been observed, particularly after cytostatic treatment with alkylating agents and topoisomerase inhibitors as treatment for primary tumors, most often for AML, myelodysplastic syndromes, or breast cancer.

CONGENITAL DISORDERS

Patients with some rare congenital chromosomal abnormalities have a higher risk of development of acute leukemia (e.g., Klinefelter's syndrome, Fanconi's anemia, Bloom's syndrome, ataxia-telangiectasia, and neurofibromatosis). Those with Down's syndrome have a twentyfold increased incidence of leukemia; ALL is increased in childhood and AML at an older age.

INFECTIOUS AGENTS

No direct evidence implicates viruses as a major cause of human acute leukemia. However, viruses are involved in the pathogenesis of two lymphoid neoplasias. In the endemic African type of Burkitt's lymphoma, the Epstein-Barr virus, a DNA virus of the herpes family, has been implicated as a potential causative agent (see [Chap. 194](#)). Endemic infection with human T-cell leukemia virus I in Japan and the Caribbean has been shown to be an etiologic agent for rare cases of adult T-cell leukemia/lymphoma (see [Chap. 201](#)).

DIAGNOSIS AND CLASSIFICATION

The diagnosis of acute leukemia is first made by examination of the peripheral blood and bone marrow. For further classification of the leukemic blast cells, cytochemical stains, immunologic markers, and cytogenetic and molecular analysis are required. The immunologic markers are still the major criteria to subdivide into B-cell lineage or T-cell lineage ALL leukemias.

PERIPHERAL BLOOD

Peripheral blood counts and a differential count from a Wright-Giemsa-stained blood smear are essential at the time of presentation. The white blood cell (WBC) count in ~40% of ALL patients is reduced or normal ([Table 106-1](#)). Only 16% of patients have a WBC above

TABLE 106-1 Laboratory Values at Diagnosis of Acute Lymphoblastic Leukemia (ALL)		
NO.		ALL
Initial white blood cell count ($\times 10^9/L$)	<10	41%
	10–50	31%
	>50–100	28%
	>100	16%
Neutrophils ($\times 10^9/L$)	<50–100	12%
	<100,000	16%
Platelets ($\times 10^9/L$)	<20	22%
	21–40	22%
	41–100	29%
	>100	27%
Hemoglobin (g/dL)	<7	20%
	7–9	33%
	>9	47%
Leukemic blasts in peripheral blood	0%	8%
	25–75%	34%
	>75%	36%
Leukemic blasts in bone marrow	<50%	4%
	51–90%	25%
	>90%	71%

Source: Data from three consecutive German Multicenter Trials for Adult ALL (GMALL).

$100 \times 10^9/L$. It is noteworthy that in 8% of ALL patients, no circulating leukemic blast cells were observed. Thus, in the frequently used automatic blood cell counting, the diagnosis may not be detected.

Peripheral blood characteristically shows anemia, thrombocytopenia, and neutropenia. Nearly one-third of patients have hemoglobin levels $<7–8$ g/dL. A platelet count below the critical number of $20 \times 10^9/L$ and neutropenia (neutrophils $<0.5 \times 10^9/L$), which is associated with a higher risk of infection, are each noted in one-fifth of adults with ALL.

BONE MARROW EXAMINATION

Bone marrow aspirates/biopsies are important to assess immunologic, cytogenetic, and genomic markers. Direct smears from the bone marrow are essential to confirm the diagnosis of acute leukemia and to distinguish between AML and ALL. The bone marrow is usually heavily packed with leukemic blast cells with $>90\%$ in $\sim 70\%$ of patients, and thus, the normal hemopoietic elements are greatly reduced or absent. A biopsy of the bone marrow will further demonstrate marked hypercellularity with replacement of fat spaces, normal elements, and occasionally increased fibrosis.

LUMBAR PUNCTURE

The examination of the cerebrospinal fluid is an essential routine diagnostic measure for ALL. Central nervous system (CNS) leukemia is diagnosed if ≥ 5 cells/ μL or leukemic blast cells were observed by morphology in cerebrospinal fluid. Opinions differ as to when the first lumbar puncture should be done—i.e., either delay lumbar puncture until remission is achieved to avoid seeding of the CNS with leukemic blast cells from the peripheral blood during the spinal tap, or perform the lumbar puncture before treatment starts, since early recognition of CNS disease will lead to immediate CNS-specific therapy. Lumbar puncture is restricted to patients with an adequate platelet count ($>20 \times 10^9/L$) and without manifest clinical hemorrhages. To eliminate potentially transferred blast cells, patients should receive intrathecal methotrexate at the first lumbar puncture.

MORPHOLOGIC SUBTYPES IN ALL

The French-American-British (FAB) classification distinguished three subgroups. L1 and L2 morphology has no clinical consequences. Only the L3 morphology, observed in up to 5% of adult patients, is indicative for mature B-cell lineage ALL (B-ALL) (see Chap. 62).

IMMUNOLOGIC SUBTYPES

A series of monoclonal antibodies is employed to identify antigens expressed on the surface of leukemic cells, corresponding to the pathways of normal B-cell differentiation (see Fig. 108-2). The aim of the immunologic classification is to subdivide ALLs according to the presence or absence of B-cell or T-cell markers. A marker is considered positive if $>20\%$ of the cells are stained with the monoclonal antibody.

There are different immunologic classifications, such as that of the European Group for the Immunological Characterization of Leukemias (EGIL), with clear therapeutic implications. Table 106-2 gives a simplified correlation of immunologic subtypes, cytogenetics and molecular aberrations, and clinical characteristics.

B-Cell Lineage ALL (B-ALL) More than 70% of adult ALLs are of B-cell origin, and the most frequent immunologic subtype, common ALL, is characterized by the presence of the ALL antigen CD10 without markers of relatively mature B cells such as cytoplasmic or surface membrane immunoglobulins. Pre-B-ALL (early B-ALL) is characterized by the expression of cytoplasmic immunoglobulin, which is negative in common ALL, but otherwise is identical with respect to all other cell markers. Pro-B-ALL corresponds to early B-cell differentiation and was formerly termed non-T-, non-B-ALL or null ALL because neither T-cell nor B-cell features could be demonstrated. This subtype is HLA-DR, terminal deoxynucleotidyl transferase, and CD19 positive and composes $\sim 12\%$ of adult ALLs. Mature B-ALL is seen in 3–4% of adults and is also known as Burkitt's leukemia. In mature B-ALL, blast cells express surface antigens of mature B cells, including the sIgM.

T-Cell Lineage ALL (T-ALL) Approximately 25% of adult ALLs are of T-cell lineage. All cases express the T-cell antigen CD7 and cytoplasmic CD3 (CyCD3) or surface CD3. According to their phase of T-cell differentiation, they may express other T-cell antigens (e.g., the E-rosette receptor CD2 and/or the cortical thymocyte antigen CD1a). Early pro/pre-T-ALL (also termed early T precursor ALL [ETP-ALL]), cortical or thymic T-ALL, and mature T-ALL can be distinguished with these markers. ETP-ALL is characterized by lack of CD1a and CD8, weak CD5 expression, and at least one myeloid/stem cell marker.

Biphenotypic or Mixed Leukemias Biphenotypic leukemias are defined as those expressing markers of both lymphoid and myeloid lineages on the same leukemic cells. Bilineage leukemias are those with two populations of blast cells with either lymphoid or myeloid antigens. It is not clear whether these patients should receive an ALL or AML treatment protocol. In pediatric studies, starting with a pediatric ALL protocol seemed preferable, which was then followed by AML consolidation elements.

CYTOGENETIC AND MOLECULAR ANALYSIS

Cytogenetic and molecular analyses should be performed in all cases in ALL. They are important to define ALL subtypes, can identify independent prognostic markers of disease-free survival, and may determine specific targeted therapies.

The diagnostic techniques for ALL are standard cytogenetics, fluorescence in situ hybridization, and reverse transcriptase polymerase chain reaction. These methods allow the detection of Ph+ ALL, with the chromosomal translocation t(9;22)(q34;q11) and the detection of the corresponding *BCR-ABL1* gene rearrangement. Further ALL entities that have been identified are t(4;11)(q21;q23)/*MLL-AFA4*, abn11q23/*MLL*, and t(1;19)(q23;p13)/*PBX-E2A*.

Gene expression profiling, single nucleotide polymorphism array analysis, array-comparative genomic hybridization, and next-generation sequencing recognize the newly defined ALL entities: ETP-ALL and Ph-like ALL.

Ph-like ALL, also known as *BCR-ABL1*-like ALL, is characterized by genetic lesions similar to Ph+ ALL, associated with *IKZF1* (Ikaros) gene deletion, *CLRF2* (gene for cytokine-like receptor-2) overexpression, and tyrosine kinase activating rearrangements involving *ABL1*, *JAK2*, *PDGFRB*, and several other genes; however, it is *BCR-ABL1* negative. The frequency is 10% in children and 25–30% in young adults but does not increase further with age like Ph+ ALL. Treatment based on the underlying genetic lesion with *BCR-ABL* inhibitors

TABLE 106-2 Immunologic, Cytogenetic, Molecular, and Clinical Characteristics of Adult Acute Lymphoblastic Leukemia (ALL)

SUBTYPES	MARKER	INCIDENCE	FREQUENT CYTOGENETIC ABERRATIONS	GENETIC ABERRATIONS AND FUSION TRANSCRIPTS	CLINICAL CHARACTERISTICS	RELAPSE KINETICS AND LOCALIZATION
B-lineage ALL (B-ALL)	HLA-DR+, TdT+, CD19+, and/or CD79a+, and/or CD22+	76%				
Pro B-ALL	No additional differentiation markers Frequent myeloid coexpression (>50%) CD10-	12%	t(4;11) (q21;q23)	70% <i>ALL1-AF4</i> (20% Flt3 in MLL+)	High WBC (>100,000/ μ L) (26%)	Mainly BM (>90%)
Common ALL	CD10+	49%	t(9;22)(q34;q11) del(6q)	33% <i>BCR-ABL</i> with 54% <i>IKZF1</i> del >25% <i>CDKN2A/B</i>	Higher age >50 years (24%)	Mainly BM (>90%) Prolonged relapse kinetics (up to 5–7 years)
Pre-B-ALL	CD10 \pm , sIg \pm	11%	t(9;22)(q34;q11) t(1;19)(q23;p13)	4% t(1;19)/ <i>PBX-E2A</i>		
Mature B-ALL	CD10 \pm , sIg \pm	4%	t(8;14)(q24;q32) t(2;8)(p12;q24) t(8;22)(q24;q11)		Higher age >55 years (27%) Frequent organ involvement (32%) and CNS involvement (13%)	Frequent CNS (10%) Short relapse kinetics (up to 1–1.5 years)
T-lineage ALL (T-ALL)	cyCD3 or sCD3	24%	t(10;14)(q24;q11) t(11;14)(p13;q11)	50% <i>NOTCH1B</i> 33% <i>HOX11b</i> 5% <i>HOX11L2b</i> 4% <i>NUP213-ABL1</i>	Younger age (90% <50 years) Frequent mediastinal tumors (60%) Frequent CNS involvement (8%) High WBC (>50/ μ L) (46%)	Frequent CNS (up to 10%) Extramedullary (6%) Intermediate relapse kinetics (up to 3–4 years) When relapsed, fast progression
Early Pro/ Pre T-ALL	No additional differentiation markers, mostly CD2–CD1a+, sCD3 \pm	6%				
Cortical T-ALL		12%				
Mature T-ALL	sCD3+, CD1a–	6%				

Abbreviations: BM, bone marrow; CNS, central nervous system; WBC, white blood cells.

(e.g., dasatinib) or JAK2 inhibitors (e.g., ruxolitinib) has so far had limited success in adults.

MINIMAL RESIDUAL DISEASE

Minimal residual disease (MRD) is the detection of residual leukemic cells that are not recognizable by light microscopy. Methods for determining MRD are based on the detection of leukemia-specific aberrant immunophenotypes by flow cytometry, the evaluation of leukemia-specific rearranged immunoglobulin or T-cell receptor sequences by real-time quantitative polymerase chain reaction, or the detection of fusion genes associated with chromosomal abnormalities (e.g., *BCR-ABL*, *MLL-AF4*). The detection limit with these methods is 10^{-3} – 10^{-5} (0.1–0.001%). With new techniques such as next-generation sequencing (NGS) or digital droplet polymerase chain reaction (ddPCR), the sensitivity may increase to 10^{-5} – 10^{-6} . The phenotypic aberrations are unique to each patient with ALL and can be detected in up to 95% of individuals. Collection of bone marrow at diagnosis for identification of patients' individual markers is essential for follow-up of MRD.

MOLECULAR RESPONSE AFTER INDUCTION THERAPY AND IMPACT ON OUTCOME

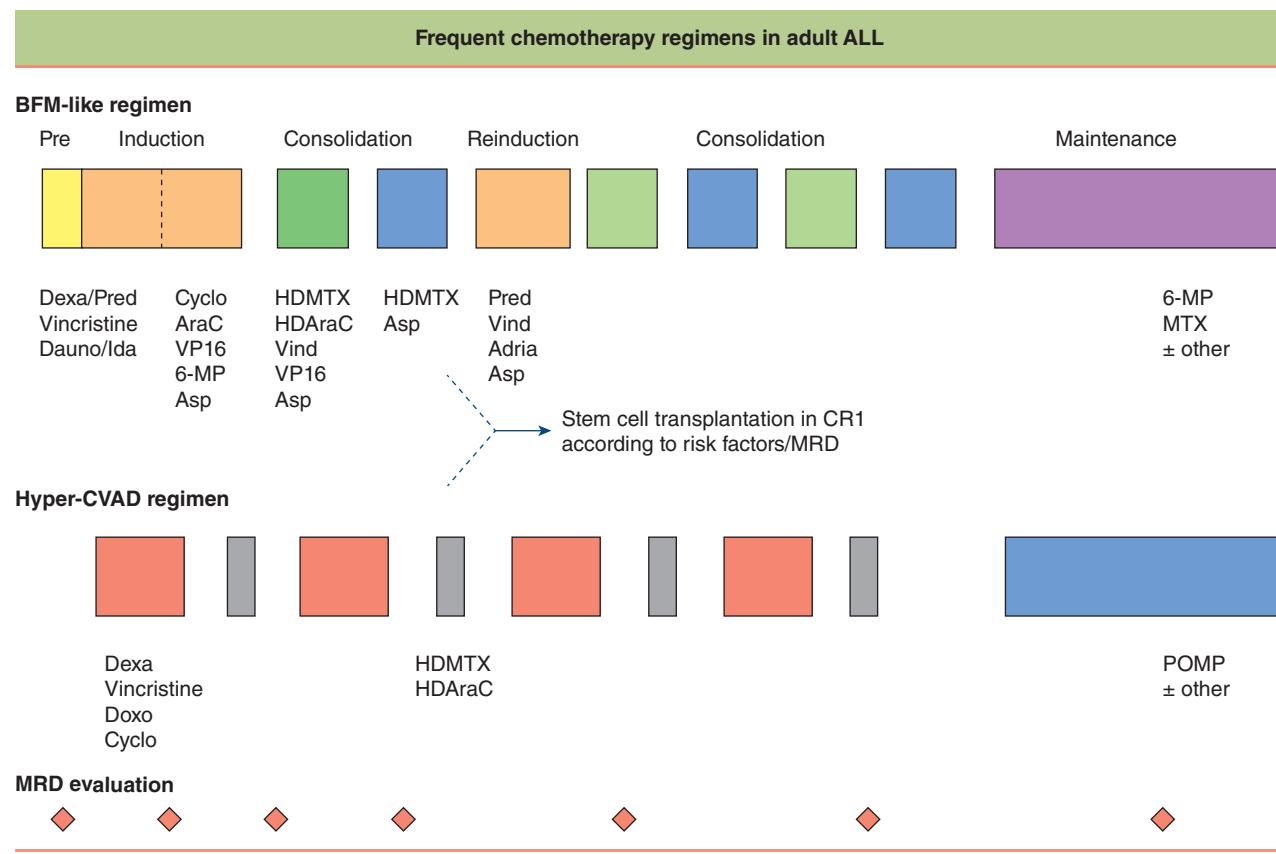
Achievement of molecular complete response/molecular remission is the most relevant independent prognostic factor for disease-free survival and overall survival in pediatric and adult ALL (Table 106-3). Patients with molecular complete remission after induction therapy had significantly superior outcomes in several studies, with a disease-free survival rate of ~70% compared to <40% for MRD-positive patients. Patients with molecular failure after induction should proceed to a targeted therapy to reduce the tumor load, followed by allogeneic stem cell transplantation (SCT), if possible.

PROGNOSTIC FACTORS, RISK STRATIFICATION, AND MRD

The aim of identification of prognostic parameters at diagnosis, which include age, white blood cell count, immunophenotype, and cytogenetic and genetic aberrations, is to stratify patients into risk groups: standard-risk patients are patients without any risk factor, and high-risk patients are those with one or more risk factors. High-risk patients are most often candidates for SCT in first complete remission (CR). MRD is thus the most important prognostic factor during therapy (Fig. 106-1); 20–30% of adult ALL patients who are MRD negative after induction will relapse. Potential reasons include loss of sensitivity, evolution of leukemic subclones, and extramedullary origin of disease. If the MRD status of a patient is not available, risk stratification should rely on clinical and laboratory risk factors evaluated at diagnosis.

TABLE 106-3 Response Parameters According to Minimal Residual Disease (MRD)

TERMINOLOGY	DEFINITION
Complete hematologic remission (CHR)	Leukemic cells not detectable by light microscopy (<5% blast cells in bone marrow [BM])
Complete molecular remission/MRD negativity	Patient in complete remission, MRD not detectable, $\leq 0.01\% = \leq 1$ leukemia cell in 10,000 BM cells
Molecular failure/MRD positivity	Patient in complete hematologic remission, but not in molecular complete remission $>0.01\%$
Molecular relapse/MRD positivity	Patient still in complete remission, had prior molecular complete remission, leukemic blast cells in BM not detectable (<5%)
Hematologic relapse	>5% blast cells in BM/blood



- Prophylactic CNS treatment; intrathecal monotherapy; MTX or intrathecal triple MTX, AraC, Dexa/Pred, +/- cranial irradiation (24 Gy)
- MRD evaluation; material collection at diagnosis, evaluation after induction 1, induction 2, and consolidation 1, and then every 3 months
- Rituximab in B-lineage, nelarabine in T-lineage
- Maintenance therapy, ~2 years in all subtypes (except Burkitt's)

FIGURE 106-1 A schematic treatment algorithm in acute lymphoblastic leukemia (ALL). 6-MP, 6-mercaptopurine; Adria, Adriamycin (doxorubicin); AraC, cytarabine; Asp, asparaginase; BFM, Berlin-Frankfurt-Münster; CNS, central nervous system; CR1, first complete remission; Cyclo, cyclophosphamide; Dauno, daunorubicin; Dexa, dexamethasone; Doxo, doxorubicin; HD, high-dose; Hyper-CVAD, hyperfractionated cyclophosphamide, vincristine, doxorubicin, and dexamethasone; Ida, idarubicin; MRD, minimal residual disease; MTX, methotrexate; POMP, mercaptopurine, vincristine, methotrexate, and prednisolone; Pred, prednisolone; Vind, vindesine; VP16, etoposide.

TREATMENT PRINCIPLES

Treatment of ALL consists usually of pre-phase therapy, induction therapy, consolidation cycles, and maintenance treatment. Treatment should start immediately when the diagnosis of ALL is established.

Pre-Phase Therapy Pre-phase therapy consisting of glucocorticoids (prednisone 20–60 mg/d or dexamethasone 6–16 mg/d, both IV or PO) alone or in combination with another drug (e.g., vincristine, cyclophosphamide) is usually given for ~5–7 days. It allows safe tumor reduction to avoid tumor lysis syndrome, to initiate supportive therapy, such as substitution of platelets/erythrocytes, or to treat infections. The time required for pre-phase therapy will also allow time to obtain results of the diagnostic workup (e.g., cytogenetics, molecular genetics).

Induction Therapy The goal of induction therapy is the achievement of a CR or, even better, a molecular CR. With current regimens, the CR rate has increased to 80–90% and is higher for standard-risk patients (>90%) and lower for high-risk patients (~60%).

Induction regimens are centered around vincristine, glucocorticoids, and anthracyclines with or without cyclophosphamide or cytarabine. L-Asparaginase is the only ALL-specific drug and is now more intensively used in adults. Pegylated asparaginase has the advantage of a significantly longer period of asparagine depletion. Dexamethasone is often preferred to prednisone because it penetrates the blood-brain barrier and also acts on resting leukemic blast cells.

Two chemotherapy regimens are widespread (Fig. 106-1). One is patterned after the pediatric BFM (Berlin-Frankfurt-Münster) protocol, which is mostly used in European adult ALL trials. Another

approach is to repeat two different alternating intensive chemotherapy cycles, identical for induction and consolidation, for eight cycles, such as Hyper-CVAD (hyperfractionated cyclophosphamide, vincristine, doxorubicin, and dexamethasone) protocol, which is preferentially used in the United States but also in many other parts of the world.

Postremission Consolidation Usual protocols use six to eight courses and often contain systemic high-dose (HD) therapy to reach sufficient drug levels in sanctuary sites such as the CNS. Most often HD methotrexate ($1-1.5 \text{ g/m}^2$ and up to $3-5 \text{ g/m}^2$) and/or HD cytarabine (4–12 doses at $1-3 \text{ g/m}^2$) are administered.

Maintenance Therapy Maintenance therapy, a strategy transferred from childhood ALL, is mandatory. It consists of 6-mercaptopurine and methotrexate plus intrathecal therapy. The potential effect of further intensification cycles during maintenance remains unclear. The duration of maintenance therapy for T-ALL and B-ALL is 2–2.5 years, except for Burkitt's leukemia, for which it is not required. In Ph+ ALL, patients also require maintenance therapy that should include a tyrosine kinase inhibitor (TKI), most likely the TKI that has been used during induction and consolidation therapy. It is also standard to give a TKI after allogeneic SCT. The duration of maintenance therapy with a TKI is also 2–2.5 years and should be guided by MRD evaluation. TKI use is often interrupted or switched to another TKI if toxicity occurs.

TREATMENT OF ALL PATIENTS ACCORDING TO AGE

The outcome of ALL is strictly related to the age of a patient, with cure rates of ~90% in children, decreasing to <10% in elderly or frail

TABLE 106-4 Best Results in Recent Studies for Adult Acute Lymphoblastic Leukemia (ALL)

Subtype	Treatment	Overall Survival
Burkitt's leukemia	Short intensive chemotherapy + rituximab; no SCT; no maintenance	80–90%
B-lineage ALL, Ph-		
AYA 15–35/45 years	Pediatric inspired, few/no SCT	≥70–80%
Adults 45–55 years	Intensive chemotherapy +/- SCT	50–60%
Elderly 55–70 years	Less intensive chemotherapy + immunotherapy	~30%
Frail >70/75 years	Various	≤10%
B-lineage ALL, Ph+		
Ph BCR-ABL	Intensive chemotherapy + TKI +/- SCT	60–70%
Ph-like ALL	Chemotherapy + dasatinib/JAK inhibitors	≤50%
T-lineage ALL		
Early (ETP)	Intensive chemotherapy + nelarabine + SCT	40–50%
Cortical/thymic	Intensive chemotherapy + nelarabine, no SCT	70%
Mature	Intensive chemotherapy + nelarabine + SCT	30–50%

Abbreviations: AYA, adolescent and young adult; ETP, early T precursor; Ph, Philadelphia chromosome; SCT, stem cell transplantation; TKI, tyrosine kinase inhibitor.

patients. Thus, age-adapted protocols have emerged, where the age limits are directed by the hematologic and nonhematologic toxicities. **Table 106-4** provides a summary of the best results obtained in adult ALL according to ALL subtype, age, and treatment. Molecular CRs are often durable. The major risk of relapse is in the first 2 years; thereafter, relapse is much less likely.

PROPHYLAXIS AND TREATMENT OF CENTRAL NERVOUS SYSTEM LEUKEMIA

Prophylactic CNS therapy in ALL is essential in order to prevent CNS leukemia and to avoid spread of leukemic cells from the CNS back to the periphery. Treatment options include intrathecal therapy, systemic HD chemotherapy, and cranial radiation therapy (CRT). Intrathecal therapy mostly consists of methotrexate as a single drug or in combination with cytosine arabinoside (AC) with or without glucocorticoids. The route of intrathecal therapy application is generally lumbar puncture. Systemic HD chemotherapy may comprise HDAC or HD methotrexate since both drugs reach cytotoxic drug levels in the CSF and show effectiveness in overt CNS leukemia. CRT (18–24 Gy in 12 fractions over 16 days) is also effective as preventive treatment of CNS leukemia. Using combined modalities for CNS prophylaxis, the CNS relapse rate has decreased to 2–5%.

Particular attention to CNS prophylaxis is required for targeted therapies. In Ph+ ALL, not all TKIs cross the blood-brain barrier equally. Dasatinib and probably ponatinib do cross the blood-brain barrier, whereas imatinib and nilotinib do not. In addition to immunotherapy, intrathecal therapy is required because most antibodies do not enter the CNS.

CNS involvement at diagnosis is observed in 5–10% of adult patients and is higher in mature B-ALL (up to 10–15%) and T-ALL (up to 10%). Treatment consists of the standard chemotherapy with additional intrathecal applications 3–5 times per week until blast cells are cleared in the spinal fluid. Patients with initial CNS involvement have a similar overall survival as CNS-negative patients.

Relapse in CNS is usually accompanied by bone marrow involvement, and if blast cells are not seen morphologically, MRD as a sign of discrete infiltration is positive in nearly all cases. CNS relapse requires local as well as systemic therapy. The outcome after CNS relapse is

dismal, and salvage chemotherapy followed by allogeneic SCT is the most effective option. Chimeric antigen receptor (CAR) T cells (most often targeting CD19) can cross the blood-brain barrier and achieve CRs in patients with CNS relapse.

STEM CELL TRANSPLANTATION

SCT is an essential part of the treatment strategy for adult ALL. Peripheral blood cells are increasingly being used as a stem cell source, instead of bone marrow. In addition, a shift from sibling stem cell donors to matched unrelated donors or haploidentical transplants from relatives has occurred. Indications for SCT in first CR are controversial. However, in most studies, SCT is recommended for high-risk patients defined either by conventional prognostic factors or by MRD positivity. High-risk patients transplanted in first CR have a survival rate of 50% or greater; decreasing transplant-related mortality from 20–30% to 10–15% has contributed substantially to better outcomes. For standard-risk patients with sustained molecular remission, allogeneic SCT in first CR is not recommended. Autologous SCT should be restricted to MRD-negative patients, BCR-ABL-negative patients, Ph+ patients, and older patients because it is less toxic but associated with a substantially higher relapse rate. For all relapsed adult ALL patients, an allogeneic SCT is thus far the only curative option.

PEDIATRIC-INSPIRED THERAPIES FOR ADOLESCENTS AND YOUNG ADULTS

The principle of pediatric-inspired therapies is to have higher doses and more applications of ALL-specific drugs such as glucocorticoids, vincristine, and L-asparaginase and fewer myeloablative anthracyclines or alkylating agents, with strict adherence to time-dose intensity, thereby reducing the role of SCT. The overall survival rates for AYAs are 70–80%.

ADULT ALL

The treatment results for adult ALL patients have greatly improved with more intensive chemotherapy, optimized SCT, and better supportive care. In several recent multicenter prospective trials, the overall survival rate for standard-risk patients was >70% with chemotherapy alone, and for high-risk patients, the overall survival rate has increased from 20–30% to >50%.

ELDERLY ALL

Palliative treatment regimens for elderly patients have failed, with CR rates of ~40%, a high early death rate of 24%, and a poor overall survival of only a few months. Intensive chemotherapy has also failed, with a higher CR rate of 56%, but still an early death rate of 23%, and only moderate improvement of overall survival to 14 months. Specific elderly ALL protocols with less intensive therapy based on glucocorticoids, vincristine, and asparaginase, largely avoiding anthracyclines and alkylating agents, have improved outcomes. The early treatment-related death rate decreased to <10%, CR rates improved to ~90%, and overall survival of ~30 months was noted.

Frail patients above the age of 70–75 years have very poor survival of <10%. Hopefully, this will improve with ongoing targeted therapies with either TKIs in Ph+ ALL or immunotherapies.

TARGETED THERAPIES

Substantial progress in adult ALL has been made in the past decade by the introduction of new targeted therapies, including TKIs and immunotherapeutic approaches (**Table 106-5**).

TYROSINE KINASE INHIBITORS IN PHILADELPHIA-POSITIVE ALL

Patients with Ph+ ALL constitute ~25% of adult B-ALL patients, with the frequency increasing to ~50% among elderly patients. In the pre-imatinib era, CR rates were 60–70%; survival with chemotherapy was ~10%, and after allogeneic SCT, it was ~30%. With the first-generation TKI imatinib, CR rates increased to 80–90%, the rate of BCR-ABL negativity increased from 5 to 50%, and the 5- to 10-year overall survival improved to 50–70%.

Faster and deeper molecular responses are achieved with second-generation TKIs (dasatinib, nilotinib), and these responses apparently

TABLE 106-5 Targeted Therapies in Adult Acute Lymphoblastic Leukemia (ALL)**Tyrosine Kinase Inhibitors (TKIs)****Ph/BCR-ABL+ ALL**

TKIs

Imatinib, dasatinib, nilotinib, bosutinib, ponatinib

Ph/BCR-ABL-like ALL

ABL1, ABL2: dasatinib; JAK2: ruxolitinib

Immunologic Approaches**Antibodies directed leukemia surface antigens**

Monovalent antibodies

Bivalent antibodies against the tumor and CD3 (e.g., blinatumomab)

Adoptive cellular therapy

T cells engineered to kill leukemic cells

Checkpoint Inhibitors

translate into a survival benefit. The third-generation TKI ponatinib is also effective in tumors bearing mutations (particularly T315I) that convey resistance to earlier-generation TKIs.

Treating adult Ph+ ALL with an allogeneic SCT in first CR is still a good treatment option for adult patients, with a 5-year overall survival of 60–70%. In elderly patients, when low-intensity chemotherapy was combined with dasatinib, the CR rate was >90%. In a next step, by combining mini-chemotherapy with a TKI and adding immunotherapy with inotuzumab (an anti-CD22 antibody), the CR rate was >90% and the overall survival improved further. A pilot experience with a chemotherapy-free regimen composed of dexamethasone, the TKI dasatinib, and the bispecific antibody blinatumomab (anti-CD19 and anti-CD3) demonstrated a CR rate of 98% and 2-year overall and disease-free survival rates of 95% and 88%, respectively. Blinatumomab eliminates Ph+ leukemic cells with resistant mutations.

IMMUNOTHERAPEUTIC APPROACHES

Treatments involving monoclonal antibodies or activated T cells are currently changing the treatment paradigm of ALL. The prerequisite is that B-lineage blast cells express a variety of specific antigens, such as CD19, CD20, and CD22 (**Table 106-6**) that are targetable with a wide variety of monoclonal antibodies. A new treatment principle is the activation of the patient's T cells to destroy their CD19+ leukemic blasts.

Anti-CD20 The anti-CD20 monoclonal antibody rituximab has improved the outcome of patients with de novo Burkitt's leukemia/

lymphoma. With repeated short cycles of intensive chemotherapy combined with rituximab, the overall survival increased to >80%. Rituximab is now included in most B-ALL regimens and is given at the usual dose of 375 mg/m² on day -1 before chemotherapy for at least eight or more cycles. This leads to a significant increase in MRD negativity and improved survival.

Anti-CD22 Monoclonal antibodies directed against CD22 are linked to cytotoxic agents, such as calicheamicin (inotuzumab ozogamicin), or to plant or bacterial toxins (epratuzumab). In a randomized trial of relapsed or refractory ALL patients, the CR rate was 66% and significantly superior to the CR rate with standard chemotherapy. Inotuzumab is now included in first-line therapy for Ph+ and Ph- patients.

Anti-CD19 Targeting CD19 is of great interest because this antigen is highly expressed in all B-lineage cells, most likely including early lymphoid precursor cells. A new promising approach is the bispecific antibody blinatumomab, which combines single-chain antibodies to CD19 and CD3, such that T cells lyse the CD19-bearing B cells.

Blinatumomab is particularly effective in MRD-positive patients, with a 70–80% conversion to MRD negativity, translating into improved overall survival; ~25% of MRD-negative patients survived without any further treatment. Blinatumomab has also moved to frontline therapy.

CAR-T Cells The adoptive transfer of CAR-modified T cells directed against CD19 is a promising approach for the treatment of CD19+ childhood or adult ALL. In the first three larger studies in adults with relapsed or refractory ALL, the CR rate ranged from 67 to 91% with MRD negativity in 60–81% of the patients who achieved CR. Overall survival is 50% or more at ≥2 years, which is remarkable for these heavily pretreated patients. CAR-T cells are also effective in CNS leukemia and in other extramedullary sites. CAR-T cell therapy in relapsed or refractory ALL was first considered as a bridge to allogeneic SCT, applied in 10–50% of patients, but the necessity for an allogeneic SCT after CAR-T cells is unclear. CAR-T cell therapies are also moving to the frontline. CD19-negative relapses after CAR-T cell therapy or blinatumomab due to downregulation of CD19 expression are a relevant obstacle.

Toxicities of Immunotherapies The anti-CD22 agent inotuzumab ozogamicin is associated with hepatotoxicity, including veno-occlusive disease, particularly after allogeneic SCT, but can be managed by reduced dosing and limitation of cycles. For anti-CD19 therapies, cytokine release syndrome and severe neurotoxicity are the most prominent toxicities and often require intensive care unit care (more so after CAR-T cells than blinatumomab). Management of these complications has improved with early recognition. Because toxic death after immunotherapies is very low compared to intensive chemotherapy or allogeneic SCT, immunotherapies are now increasingly included in frontline therapy.

TREATMENT OF T-ALL

Immunotherapy for T-ALL is still not available and intensive chemotherapy is still the mainstay in combination with the T cell-specific drug nelarabine. Currently, γ-secretase targeting NOTCH1, checkpoint inhibitors such as bortezomib and venetoclax, and HDAC inhibitors are being explored.

CONCLUSION AND FUTURE DIRECTIONS

Cytogenetic and molecular analysis at diagnosis allows identification of ALL subentities, requiring different treatment options. Evaluation of MRD is the most important parameter for treatment decisions. The greatest progress has been achieved by targeted therapies, such as TKIs for Ph+ ALL and new immunotherapeutic approaches. This will lead to further improved outcome of adult ALL patients, 50% of whom are already surviving 5–10 years and are most likely cured. New options and advances, such as low-intensity chemotherapy, reduction of SCT, incorporation of targeted therapies, and reduction of toxicities, will improve the quality of life of patients and lead to individualized approaches for each patient.

TABLE 106-6 Expression of Antigens in B-Cell Lineage Acute Lymphoblastic Leukemia (ALL) for Potential Antibody Therapy

SURFACE ANTIGEN	ALL SUBTYPES	EXPRESSION ON LBC ^a	MONOCLONAL ANTIBODY
CD20	Burkitt's lymphoma/leukemia	86–100%	Rituximab
	B-precursor	30–40%	Ofatumumab
CD22	B-precursor	93–98%	Inotuzumab
	Mature B-ALL	~100%	Epratuzumab Moxetumomab pasudotox
CD19	B-precursor	95–<100%	T cell-activating therapies
	Mature B-ALL	94–<100%	Blinatumomab Bispecific CD3/CD19 Chimeric antigen receptor modified T cells (CAR T cells)

^aDefined as ≥20% positive blast cells.

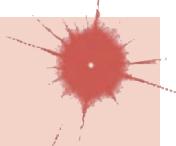
Abbreviation: LBC, leukemic blast count.

Source: Republished with permission of American Society of Hematology, from D Hoelzer: Novel antibody-based therapies for acute lymphoblastic leukemia. 2011: 243, 2011; permission conveyed through Copyright Clearance Center, Inc.

107

Chronic Lymphocytic Leukemia

Jennifer A. Woyach, John C. Byrd



Chronic lymphocytic leukemia (CLL) is a monoclonal proliferation of mature B lymphocytes defined by an absolute number of malignant cells in the blood ($5 \times 10^9/\text{mL}$). The presence of malignant B cells under this count in the blood without nodal, spleen, or liver involvement and absent cytopenias is a precursor of this disease called *monoclonal B cell lymphocytosis* (MBL) with ~1–2% chance per year of progressing to overt CLL. CLL is a heterogeneous disease in terms of natural history, with some patients presenting asymptotically and never requiring therapy, whereas others present with symptomatic disease, require multiple lines of therapy, and eventually die of their disease. Over the past 10–15 years, the understanding of CLL origin and biology has grown exponentially, leading first to more refined disease definition, prognostic markers, and, subsequently, introduction of novel therapies that have significantly changed the natural history of this disease. In this chapter, we review the epidemiology, biology, and management of CLL, with a focus on new knowledge that is currently changing standards of care.

EPIDEMIOLOGY

CLL is primarily a disease of older adults, with a median age at diagnosis of 71 and an age-adjusted incidence of 4.5/100,000 people in the United States. The prevalence of CLL has increased over the past decades due to improvements in therapy for this disease and also survival of older patients from other medical ailments. In 1980, the 5-year overall survival of patients was 69%, and this increased to 87.9% in 2007 and is likely even higher today. The male-to-female ratio is 2:1; however, as patients age, the ratio becomes more even, and over the age of 80, the incidence is equal between men and women. The disease is most common in Caucasians, less common in Hispanic and African Americans, and rare in the Asian population.

Unlike many other malignancies, there have been no definitive links between CLL and exposures. Indeed, CLL is one of the only types of leukemia not linked to radiation exposure. Agent Orange exposure has been implicated, and CLL is thus a service-connected condition for those who were exposed to Agent Orange in the Vietnam conflict.

FURTHER READING

- BROWN P et al: Effect of postreinduction therapy consolidation with blinatumomab vs chemotherapy on disease-free survival in children, adolescents, and young adults with first relapse of B-cell acute lymphoblastic leukemia A randomized clinical trial. JAMA 325:833, 2021.
- FOA R et al: Dasatinib-blinatumomab for Ph-positive acute lymphoblastic leukemia in adults. N Engl J Med 383:1613, 2020.
- FREY NV et al: Optimizing chimeric antigen receptor T-cell therapy for adults with acute lymphoblastic leukemia. J Clin Oncol 38:415, 2020.
- HOELZER D et al: Improved outcome of adult Burkitt lymphoma/leukemia with rituximab and chemotherapy: report of a large prospective multicenter trial. Blood 124: 3870, 2014.
- HOELZER D et al: Acute lymphoblastic leukaemia in adult patients: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 27(suppl 5):v69, 2016.
- IACOBUCCI I, MULLIGHAN C: Genetic basis of acute lymphoblastic leukemia. J Clin Oncol 35:975, 2017.
- KANTARJIAN HM et al: Inotuzumab ozogamicin versus standard therapy for acute lymphoblastic leukemia. N Engl J Med 375:740, 2016.
- ROBERTS KG et al: Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. N Engl J Med 371:1005, 2014.

CLL is one of the most familial-associated malignancies, and the first-degree relative of a CLL patient has an 8.5-fold elevated risk of developing CLL than the general population. MBL is also more common in families with two first-degree relatives having CLL, further supporting a genetic predisposition of this disease. Despite this, specific genes conferring risk in the familial setting outside of specific families have been difficult to identify. In genome-wide association studies (GWAS), ~30 single nucleotide polymorphisms have been identified, which is estimated to account for 19% of the familial risk of CLL. Genes involved in apoptosis, telomere function, B-cell receptor (BCR) activation, and B-cell differentiation have all been implicated in GWAS. Variants in shelterin complex proteins involved in telomere maintenance such as POT1 have been identified in a small number of families.

BIOLOGY AND PATHOPHYSIOLOGY

CELL OF ORIGIN

The cell of origin in CLL has not definitively been established. The morphology, immunophenotype, and gene expression pattern of CLL cells are that of a mature B cell (Fig. 107-1), and so it has been presumed that the initiating cell is a mature lymphocyte, perhaps memory B cells. However, many facets of CLL biology do not support this idea, including antigen-binding characteristics of CLL cells and the presence of stereotyped BCRs. Other possibilities include a stepwise process including a series of transforming events at various stages of B-cell development, potentially including de-differentiation of more mature cells. The self-renewing, multipotent hematopoietic stem cell (HSC) might also be the originating cell of CLL, postulated based on transplant studies in mice showing clonal leukemic cell development with different characteristics from donor leukemia after transplantation of HSCs. More work will be required to elucidate the origins of CLL.

B-CELL RECEPTOR SIGNALING IN CLL

Perhaps the most important advancement in CLL biology is the understanding of the role of BCR signaling in the disease. CLL has distinct BCR signaling as compared to normal B cells, which is characterized by low-level IgM expression, variable response to antigen stimulation, and tonic activation of antiapoptotic signaling pathways that promote tumor survival. CLL cells by gene expression profiling share many features with antigen-activated mature B cells, suggesting a role for activation of BCR signaling in the disease pathogenesis. Tissue-based microarrays have revealed upregulation of BCR pathway genes in the lymph nodes and bone marrow compared to the peripheral blood, suggesting a particular importance of this pathway in microenvironmental homing.

Fitting with the role of BCR signaling in CLL, one of the most influential prognostic factors identified in this disease is the mutational status of the immunoglobulin heavy chain variable (IGHV) region. During normal B-cell maturation, the variable regions of the

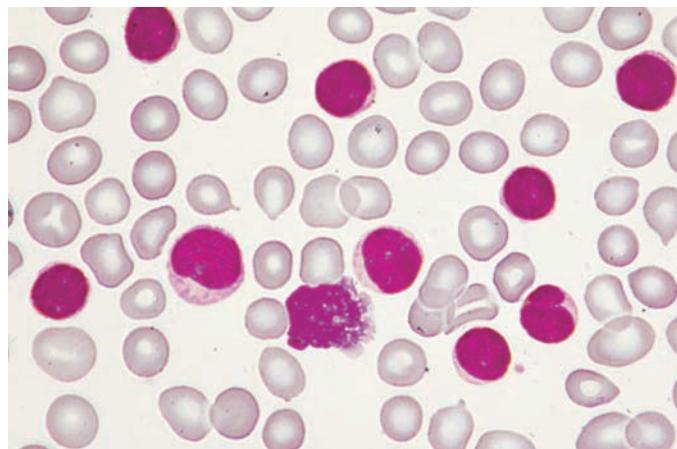


FIGURE 107-1 Chronic lymphoid leukemia in the peripheral blood. (From M Lichtman et al [eds]: Williams Hematology, 7th ed. New York, McGraw-Hill, 2005.)

immunoglobulin heavy chain undergo somatic hypermutation. In CLL, ~60% of patients have IGHV that is $\geq 2\%$ mutated from germline. This may indicate a more mature, postgerminal center progenitor, and is typically associated with a more indolent disease course. Conversely, ~40% of patients will have IGHV $< 2\%$ mutated from germline, which is associated with more rapid progression of disease and short survival prior to the era of therapeutics that target BCR. Unfavorable biologic properties including enhanced telomerase activity, overexpression of activation-induced cytidine deaminase, increased nuclear factor- κ B (NF- κ B) activity, high-risk genomic mutations (e.g., NOTCH1, SF3B1, TP53, ATM), and clonal evolution are also associated with IGHV unmutated disease.

Because IGHV sequencing was initially cumbersome to perform, a number of surrogate factors have been identified; however, none yet have been shown to be equal or superior to IGHV sequencing. The most prevalent of these surrogate markers are Zap-70 expression, ZAP-70 methylation, and surface CD38 expression. Zap-70 protein is a normal intracellular T-cell signaling protein that is aberrantly expressed in most IGHV unmutated CLL cells. CD38 is a marker that is also more highly expressed on the surface of IGHV unmutated CLL cells. Both of these prognostic factors are widely used but limited in their applicability. Zap-70 protein status is difficult to measure by flow cytometry and has low reproducibility. Measurement of methylation status of the ZAP-70 promoter is much more precise but not widely available. CD38 expression is easier to measure by flow cytometry but not as highly predictive of outcomes and can change during the course of disease.

CYTOGENETIC ABNORMALITIES

Besides IGHV mutational status, recurrent cytogenetic abnormalities are the most robust prognostic factor clinically available in CLL. These abnormalities are typically identified by fluorescent in situ hybridization (FISH) analysis; however, stimulated metaphase karyotype has a role as well. The most well-characterized abnormalities include del(13)(q14.3), trisomy 12, del(11)(q22.3), and del(17)(p13.1) (Fig. 107-2). The presence of sole del(13)(q14.3) is associated with more indolent disease, prolonged survival, and good response to traditional therapies. Usually, this abnormality is not seen on banded karyotype analysis, and when present on karyotype, it indicates a larger deletion involving the retinoblastoma gene, which negates the favorable prognosis associated with this marker. Trisomy 12 has a more intermediate prognosis. The del(11)(q23.3) results in deletion of the ATM gene and is associated with bulky lymphadenopathy and aggressive disease in young patients, with inferior prognosis, and more rapid progression to symptomatic disease. The del(17)(p13.1) results in loss of one allele of the tumor suppressor TP53 and is associated with the poorest prognosis in CLL with rapid disease progression, poor response to traditional therapies, and shorter survival. Other abnormalities have been shown to be important in smaller studies but are not routinely performed at all centers. Finally, complex karyotype (three or more abnormalities) on stimulated metaphase karyotype analysis has significant adverse impact on time to treatment and overall survival, with

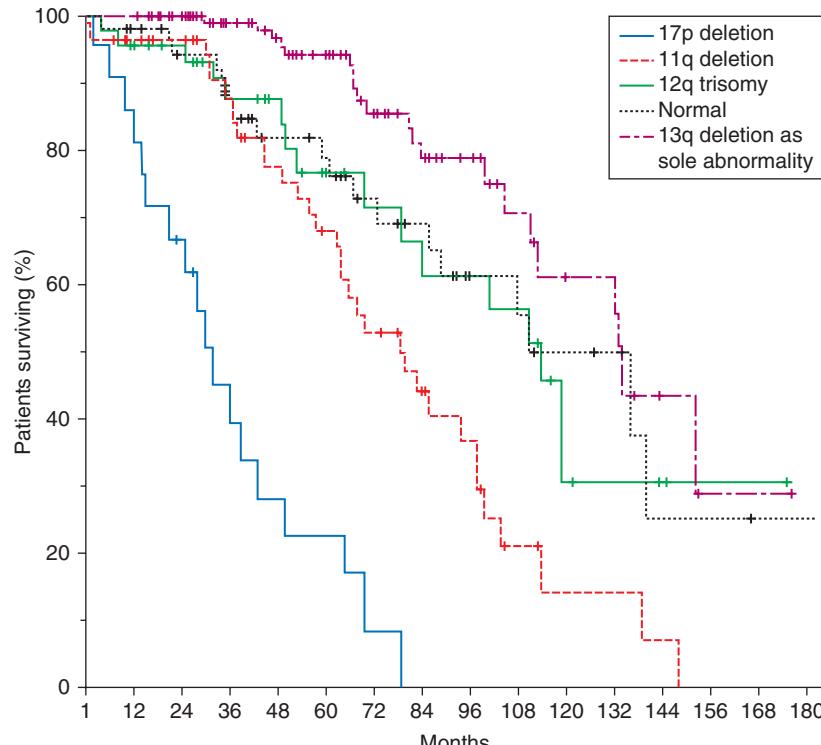
data indicating that increasing complexity is even more deleterious to response and survival.

Clonal evolution, or acquisition of cytogenetic or molecular abnormalities, is common in CLL, especially in patients with IGHV unmutated CLL. Because the tumor cytogenetics can change over time, it is recommended that FISH, with or without cytogenetics, is checked before every line of therapy, mostly to evaluate acquisition of del(17)(p13.1).

GENE MUTATIONS AND MIR ALTERATIONS

Compared with many other malignancies, the genome in CLL is relatively simple, with an average CLL genome carrying ~20 nonsynonymous alterations and ~5 structural abnormalities. And, unlike many other hematologic malignancies, there is no unifying genetic lesion, and most recurrent genetic driving mutations exist at frequencies of <5%. Whole genome and whole exome sequencing have identified the most common mutations in CLL to be in SF3B1, NOTCH1, MYD88, ATM, and TP53 (Table 107-1). Most of the identified mutations in these genes are common among different malignancies, and with the exception of MYD88, they are generally identified with much higher frequency in IGHV unmutated disease.

NOTCH1 mutations are present in ~15% of CLL patients and are commonly associated with trisomy 12. Although multiple different mutations are seen, most are located within the PEST (proline, glutamic acid, serine, and threonine) domain and result in constitutive NOTCH signaling. NOTCH1 mutations have been associated with lower sensitivity to CD20 antibody therapy and increased risk of transformation to aggressive diffuse large B-cell lymphoma (DLBCL; Richter's transformation), although its relevance in the era of targeted therapies is less clear. SF3B1 is a component of the RNA spliceosome and is mutated in 10–15% of CLL cases. Mutations appear to be associated with



No. AT Risk	17p deletion	11q deletion	12q trisomy	Normal	13q deletion as sole abnormality
17p deletion	23 18 13 8 5 4 1 0 0 0 0 0 0 0 0				
11q deletion	56 53 47 43 33 27 20 15 10 4 2 2 1 0 0				
12q trisomy	47 44 41 29 24 17 14 13 12 11 4 3 2 1 1 0				
Normal	57 51 45 37 30 27 20 17 12 11 6 5 2 2 1 1				
13q deletion as sole abnormality	117 117 106 91 80 63 45 36 24 16 12 11 3 1 1 0				

FIGURE 107-2 Outcomes among chronic lymphocytic leukemia patients with various cytogenetic abnormalities. (From H Döhner et al: Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* 343:1910, 2000. Copyright © (2000) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

TABLE 107-1 Recurrent Mutations in CLL

GENE	FREQUENCY OF MUTATIONS (%)
<i>SF3B1</i>	8–14
<i>TP53</i>	5–13
<i>NOTCH1</i>	10–13
<i>MYD88</i>	4–8
<i>ATM</i>	8–11
<i>BIRC3</i>	<5
<i>XP01</i>	<5
<i>FBXW7</i>	<5
<i>POT1</i>	<5
<i>BRAF</i>	<5
<i>EGR2</i>	<5
<i>IKZF3</i>	<5

Abbreviation: CLL, chronic lymphocytic leukemia.

intermediate-risk disease, and, functionally, *SF3B1* may be important in the response to DNA damage.

Mutations of the tumor suppressor *TP53* are found in ~5% of CLL patients with previously untreated early-stage disease and up to 40% with later stages. Seventy percent of the time, these mutations coexist with del(17)(p13.1), effectively eliminating *TP53* function. As expected, and consistent with other malignancies, *TP53* mutations are associated with a poor prognosis and expected lack of response to DNA-damaging therapies.

ATM mutations, which are heterogeneous and occur throughout the gene, occur in 10–15% of CLL patients. *ATM* mutations often coexist with del(11)(q22.3), eliminating *ATM* on the alternate allele. Similar to *TP53*, mutations in *ATM* tend to result in impaired response to DNA damage, which can reduce responsiveness to chemotherapy.

In contrast to the aforementioned mutations, those in *MYD88* tend to occur in IGHV mutated CLL and be associated with a more indolent prognosis. This gene is involved in Toll-like receptor signaling, and the most common mutation, L265P, results in constitutive activation and NF-κB activity.

Along with abnormalities in coding genes, it has become apparent that noncoding genes such as microRNAs are recurrently altered in CLL. The most common cytogenetic abnormality, del(13)(q14.3), results in loss of the miR15/16 cluster, which is important in the pathogenesis of CLL. In normal cells, miR15A/miR16A inhibits antiapoptotic gene expression (including *BCL2*, *CCND1*, *CCND3*, and *CDK6*), and this specific deletion allows for overexpression of these genes and thus increased cell survival. Loss of other miR expression such as miR-181a leads to overexpression of antiapoptotic proteins such as MCL-1 and TCL1. Overexpression of miR-155, an onco-miR associated with B cell transformation, has also been documented in the majority of CLL patients.

IMMUNOLOGY

CLL is characterized by dysregulation of the normal immune system in addition to the malignant immune cells. Besides numerical abnormalities due to bone marrow dysfunction, even in the early stages of disease, there are skewed ratios of immune cells and functional abnormalities. Innate immune system defects associated with CLL include

reduced complement proteins and activity, qualitative neutrophil defects, and functional defects of natural killer cells.

More focus has been placed on the impairments in the adaptive immune system in this disease. Within the CD4+ T-cell compartment, a qualitative defect is noted similar to chronic antigen stimulation inducing a phenotype of T-cell exhaustion typical of what is seen in chronic viral infections such as hepatitis. This has been demonstrated to lead to impaired T-cell cytotoxic capacity and reduced proliferative ability. Additionally, there are physical changes in the T-cell cytoskeleton that cause impaired immune synapse formation with antigen presenting cells. In addition to a lack of capacity to respond to pathogens, the T-cell defect in CLL also likely leads to tumor cell tolerance. During the course of the disease, the polarization of the CD4+ T cells shifts from a Th1 (cytotoxic) phenotype to a Th2 phenotype, which leads to expansion of immunosuppressive cytokines such as interleukin 10 (IL-10). Additionally, in the later stage of disease, T regulatory cells are expanded, which contributes to an immunosuppressive phenotype.

Other components of the immune microenvironment are altered as well to form a more supportive environment for the malignant cells. M2 monocytes have been shown to differentiate into a type of tumor-associated macrophage known as a *nurse-like cell* in CLL. These cells promote survival by secreting chemokines and cytokines that increase migration and activation.

The humoral immune system in CLL is also dysregulated, as is expected for a malignancy that results in very few normal B cells. Hypogammaglobulinemia is very common and affects all subclasses of immunoglobulins, occurring in ~85% of patients at some time in their disease course, and is more common as disease progresses. A correlation between low IgG and IgA and infection risk has been established, but isolated IgM reduction does not seem to be associated with excess infection risk. Also, CLL cells can secrete monoclonal IgM or IgG in a small number of cases, and this can correlate with disease progression.

CLINICAL PRESENTATION AND DIAGNOSIS OF CLL

CLINICAL PRESENTATION AND DIAGNOSIS

The presentation of CLL most commonly occurs as an incidental diagnosis made at the time of medical evaluation for another cause. In this regard, CLL is most commonly diagnosed on routine blood work demonstrating an elevated lymphocyte count in asymptomatic individuals, although some patients present with symptoms and require early therapy. When noting either an elevated total white blood cell (WBC) count with lymphocytic predominance or a normal WBC with a differential showing a lymphocytosis, the next step is to perform flow cytometry on the peripheral blood. In CLL, this will reveal the typical immunophenotype that includes the typical B cell markers CD19, CD20, CD22, and CD23; the T-cell marker CD5 (CD5 is also expressed on the B1 subset of B cells that typically has unmutated immunoglobulin and responds to antigens independent of cognate T cell help); and dim surface immunoglobulin of either kappa or lambda type (**Table 107-2**). Atypical phenotypes can be seen as well and usually can be differentiated on the basis of morphology, cytogenetics, or clinical presentation. In cases in which the clonal B cell count based on flow cytometry is $\geq 5 \times 10^9/L$, no further workup is needed to confirm the diagnosis of CLL.

Some patients will present with a small clonal proliferation of CLL cells in the peripheral blood but will also have lymphadenopathy or

TABLE 107-2 Typical Immunophenotype of CLL Compared with Other B Cell Malignancies

DISEASE	CD5	CD10	CD19	CD20	CD23	CYCLIN D1	SURFACE IG
CLL	+	-	+	+ (dim)	+	-	+ (dim)
Mantle cell lymphoma	+	-	+	+ (mod/bright)	-	+	+ (mod/bright)
Marginal zone lymphoma	-/+	-	+	+ (mod/bright)	-/+	-	+ (mod/bright)
Follicular lymphoma	-	+	+	+	+	-	

Abbreviation: CLL, chronic lymphocytic leukemia.

splenomegaly. In these cases, the likely diagnosis is small lymphocytic lymphoma (SLL), a semantic designation from CLL that denotes a primarily tissue-based disease rather than bone marrow/blood-based disease. The genetic and molecular features of SLL are identical to those of CLL. The retention of the cells in tissues may be related to the expression of a particular adhesion molecule. Thus, SLL patients are managed identically to CLL patients, and often in the later stages of disease, these patients will have blood and bone marrow involvement as well.

MONOCLONAL B-CELL LYMPHOCYTOSIS

Patients who do not meet the diagnostic criteria for CLL based on quantification of clonal B cells in the peripheral blood and who do not have associated signs of CLL including lymphadenopathy, organomegaly, or cytopenias have a disorder known as monoclonal B-cell lymphocytosis (MBL), which is now thought to precede every case of CLL. Analogous to monoclonal gammopathy of uncertain significance (MGUS) in myeloma, not all MBL progresses to CLL. MBL is initially characterized by a CLL-like immunophenotype in ~75% of cases but can also be atypical (CD23 negative or bright CD20) or CD5 negative. More relevant for prognosis is characterization by count, with low-count MBL defining those patients with $<0.5 \times 10^9$ clonal B cells/L, and high-count MBL defining those with $>0.5 \times 10^9$ but $<5 \times 10^9$ /L. Patients with low-count MBL have a negligible rate of progression to CLL, whereas those with high count progress to overt CLL at a rate of 1–2% per year, warranting continued monitoring. Population-based studies have estimated the prevalence of MBL to be up to ~12% in the general population, where it is most common in elderly men. It is especially common in first-degree relatives of CLL patients, where the frequency is ~18%.

Although the risk of MBL progression is relatively low, it has become apparent that patients still experience complications that suggest an immune dysfunction in MBL that is similar to that seen with CLL. Rates of serious infections requiring hospitalization appear to be significantly increased in MBL, similar to the rates seen in CLL. In a case-control study, patients with MBL had a 16% chance of hospitalization over a 4-year time period, compared with 18.4% in patients with newly diagnosed CLL. Secondary cancers also appear to be increased in MBL. These data suggest that monitoring for patients with MBL should focus on vaccinations and age-appropriate cancer screening, as the probability of complications appears to be higher than the risk of progression in most of these patients. Follow-up for patients with MBL can occur with the primary care physician as this does not represent a malignancy, whereas CLL is mostly comanaged with both a primary care physician and a hematologist.

COMPLICATIONS OF CLL

A significant amount of morbidity and mortality related to CLL is due to complications of the disease. In general, complications besides disease progression include infections, secondary cancers, autoimmune complications, and transformation to a more aggressive clonally related lymphoma.

■ INFECTIONS

Infections are a leading cause of both disease-related morbidity and death in patients with CLL, with ~30–50% of deaths in CLL patients attributed to infection. Owing to the immune dysfunction associated with the disease, patients are at risk for both typical and atypical infections. Besides this baseline risk of infections, most CLL therapies can increase infection risk. For many nucleoside analogue-based chemotherapy regimens used in CLL, prophylaxis for *Pneumocystis pneumonia* is indicated for at least 6 months following therapy to allow recovery of functional T cells. Viral prophylaxis is also indicated for many chemotherapy regimens and for patients with a history of varicella-zoster to diminish reactivation and morbidity from this virus.

Because of the abnormalities in cellular and humoral immunity, vaccine responses in CLL are limited in many patients, especially in the later stages of disease. In one study, one dose of 13-valent pneumococcal vaccine produced an adequate immune response in only 58%

of patients compared with 100% in age-matched controls. Despite the known limitations, vaccination against influenza and pneumococcal pneumonia is recommended in CLL. The recombinant zoster vaccine has approximately a 60% response in previously untreated CLL, is safe, and should be considered for this patient group. In contrast, live vaccines should be avoided in the setting of CLL because of the small risk of viral reactivation with an immunocompromised host.

As discussed earlier, hypogammaglobulinemia is common in CLL and can be associated with significant risk for infections, primarily of mucocutaneous etiology such as sinusitis and bronchitis. In addition, women can have frequent urinary tract infections. While administration of prophylactic intravenous immunoglobulin (IVIg) has not been shown to improve survival, it has been shown to reduce the number of minor or moderate bacterial infections and thus is indicated in patients with hypogammaglobulinemia who suffer from recurrent infections or have pulmonary bronchiectasis. We also administer at least one dose of immunoglobulin to CLL patients who develop influenza with coexisting hypogammaglobulinemia to diminish risk of postinfluenza pneumococcal pneumonia. IVIg is also indicated in patients who have been hospitalized for a serious infection and in those whose IgG level is <300 – 500 mg/dL.

■ SECONDARY MALIGNANCIES

Multiple population-based studies have shown that patients with CLL are at an elevated risk to develop other cancers, with a rate up to three times that of the general population, even in the absence of cytotoxic chemotherapy. The most common types of cancers seen in CLL are skin, prostate, and breast cancers, although other cancers are seen as well. Skin cancers are particularly common, with a rate that is 8- to 15-fold higher than in the general population, and may behave more aggressively. All CLL patients should be counseled on the use of sunscreen while outdoors and should undergo preventative skin examinations.

In one single-center study, older age at CLL diagnosis, male sex, high β_2 -microglobulin, high lactate dehydrogenase (LDH), and chronic kidney disease were associated with excess risk of other cancers; other CLL-specific risk factors have not shown association with other cancer risk.

While cancer risk is higher, no specific recommendations for increased cancer screening in CLL patients have been validated. Age- and sex-appropriate screenings should be recommended.

Conflicting data exist regarding the risk of cancers following CLL-specific therapy. Chemoimmunotherapy, in particular alkylator-containing regimens, seems to be associated with an increased risk for secondary cancers. Secondary cancers are also seen in the setting of targeted therapies. Bruton tyrosine kinase (BTK) inhibitors appear to have a secondary cancer risk similar to what is seen in the CLL population in general, but potentially a higher rate of nonmelanoma skin cancers. With short follow-up, the risk of secondary cancers appears to be slightly higher with venetoclax-based regimens than chlorambucil-based chemoimmunotherapy, and further evaluation of this trend is ongoing.

■ AUTOIMMUNE COMPLICATIONS

Autoimmune complications are frequent in CLL. Most commonly, these include autoimmune cytopenias, but autoimmune complications of other organs including glomerulonephritis, vasculitis, and neuropathies have also been reported. Of the autoimmune cytopenias, the most common is autoimmune hemolytic anemia (AIHA), which is an antibody-mediated destruction of autologous red blood cells (RBCs). Second most common is immune thrombocytopenia (ITP), which shares some features with AIHA and has a similar mechanism targeting platelets. These two syndromes may occur in isolation, occur sequentially in the same patient, or present in combination as Evan's syndrome. Pure red cell aplasia (PRCA) and autoimmune granulocytopenia (AIG) are comparatively rare and can occur alone or in combination with other autoimmune cytopenias. It is difficult to tease out whether autoimmune cytopenias lead to worse prognosis in CLL because of various complicating factors. However, it is clear that these can lead to significant morbidity, both due to the process itself and due to therapies required for management.

AIHA usually presents as an isolated anemia with an elevated reticulocyte count and features of hemolysis including elevated bilirubin and LDH and low haptoglobin. Detection of a warm IgG antibody on the surface of RBCs with a Coombs test can help solidify the diagnosis, although Coombs-negative cases can occur. Immediate therapy is almost always necessary and consists of transfusion and immunosuppression. Glucocorticoids are often used for initial therapy, although in most cases, additional treatment is needed due to either poor response or recurrence with taper of glucocorticoid dosing. Rituximab can be successful, and therapy directed toward the underlying CLL is often effective in more resistant cases. Transfusion of blood in cases of robust AIHA must be initiated with caution as transfusion reactions can be seen due to poorly matched blood, but should be pursued in those with severe, symptomatic anemia. Death from uncontrolled AIHA can occur in the absence of appropriate supportive care (**Chap. 100**).

ITP can be more difficult to diagnose as it may be difficult to differentiate from progression of disease due to the lack of laboratory tests that identify platelet destruction from this mechanism. Signs that point toward ITP include isolated thrombocytopenia and rapid decline in platelet levels in the absence of an alternative etiology. A bone marrow biopsy showing normal or increased megakaryocytes can be used to confirm the diagnosis but is often not necessary. In CLL, treatment for ITP is usually instituted when platelet levels drop to 20,000–30,000 or if evidence of bleeding complications or need for invasive procedures develops. Like AIHA, initial therapy consists of glucocorticoids and IVIg, with rituximab also being an effective method to induce long-term remissions. Also, the thrombopoietin receptor agonists romiplostim and eltrombopag are effective in secondary ITP. In many cases, ITP can be successfully treated without treating the underlying CLL. In cases in which anemia or thrombocytopenia appears, it is important to investigate the mechanism because the approach to therapy of autoimmune cytopenias in CLL differs from cytopenias due to marrow replacement (**Chap. 115**).

RICHTER'S TRANSFORMATION

One of the most devastating complications of CLL is Richter's transformation, which is transformation of CLL to an aggressive lymphoma, most commonly DLBCL. The World Health Organization also recognizes Hodgkin's lymphoma (HL) as a variant of Richter's transformation; other aggressive lymphomas are rarely identified. Some older series have included prolymphocytic transformation in this category, although this has much less prognostic impact on long-term outcome. The prevalence of Richter's transformation is difficult to estimate based on previous studies, but one prospective observational study estimated a rate of 0.5% per year for DLBCL and 0.05% per year for HL. Risk factors for development include bulky lymphadenopathy, NOTCH1 mutations, del(17)(p13.1), and a specific stereotypedIGHV usage. Lymphomas arising in the setting of CLL can either be clonally related or unrelated to the initial CLL, with prognosis significantly better for clonally unrelated lymphomas. In addition, patients with Hodgkin's transformation have improved outcome, particularly in the absence of prior fludarabine treatment. B-cell prolymphocytic leukemia (PLL) arising from CLL is currently classified as Richter's transformation as well; however, clinical features and therapy are quite different, so these two should be differentiated for therapeutic purposes.

Clinical signs of Richter's transformation include rapid progression in adenopathy, often in a specific area, and constitutional symptoms including fatigue, night sweats, fever, and weight loss. LDH is usually high. In suspected cases, the first step is ¹⁸FDG-PET/CT (fluorodeoxyglucose–positron emission tomography combined with computed tomography) scan to localize an area for biopsy. Standardized uptake values (SUVs) <5 are consistent with CLL and can rule out Richter's transformation in many cases. SUVs >5 are suspicious for Richter's transformation, with SUVs ≥10 being very concerning. Excisional biopsy is the preferred mode of diagnosis, and fine-needle aspiration should be discouraged.

Therapy for DLBCL Richter's transformation usually involves combination chemoimmunotherapy (e.g., R-CHOP [rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone], dose-adjusted

EPOCH-R [etoposide, prednisone, vincristine, cyclophosphamide, doxorubicin, and rituximab]; **Chap. 108**). Outcomes may be poor with median survivals of 6–16 months in most series for clonally related Richter's versus ~5 years for clonally unrelated. For fit patients who achieve a response with therapy, stem cell transplantation has the possibility to induce long-term remissions and should be explored. In addition, chimeric antigen receptor T-cell (CAR-T) therapy has shown promising results in small groups of patients and remains an area of active clinical investigation. Patients with Hodgkin's disease can be treated according to the algorithm for this disease, with many individuals being cured.

WORKUP OF CLL AND APPROACH TO THERAPY

WORKUP AND STAGING

Workup of a patient with a new diagnosis of CLL based on typical immunophenotyping includes a detailed history of infectious disease; family history of CLL; and careful physical examination with attention to the lymph nodes, spleen, and liver. In patients desiring to know the expected natural history of their CLL, prognostic testing using FISH and stimulated karyotype and sequencing for TP53 and IGHV mutation status can be performed. Imaging with CT scan is usually not necessary unless there are symptoms and concern for intraabdominal nodes out of proportion to peripheral nodes. Bone marrow biopsy is not undertaken until therapy is initiated except in cases of unexplained cytopenias.

STAGING

There are two widely used staging systems in CLL. The Rai staging system is used more commonly in the United States, whereas the Binet system is more commonly used in Europe. Both characterize CLL on the basis of disease bulk and marrow failure (**Table 107-3**). Both rely on physical examination and laboratory studies and do not require imaging or bone marrow analysis. While the initial staging systems could reliably predict survival in CLL, with the changes in therapy since the original description of the stages, the impact of initial stage on survival is not as clear. Cytogenetic and genomic testing can help refine outcomes of these staging tests. An international collaboration integrated both clinical and genomic staging to better predict outcome at diagnosis and time of initial treatment, which led to development of the CLL International Prognostic Index (**Table 107-4**). This index has been shown to be useful in prediction of both time to first treatment and outcome with chemoimmunotherapy. Validation in the setting of novel targeted therapies has not occurred.

CRITERIA FOR THE INITIATION OF THERAPY

Currently, a watchful waiting strategy is used for most patients with CLL, with therapy reserved for patients with symptomatic disease. This recommendation is based on multiple trials showing no survival advantage with earlier therapy, although this question continues to be a focus of active investigation.

With the exception of patients participating on early intervention studies in CLL, disease-related symptoms that require the initiation of therapy are outlined in **Table 107-5**. Except for the rare patient who

TABLE 107-3 Staging of CLL

Rai Staging System	
Low risk (stage 0)	Lymphocytosis only
Intermediate risk (stage I/II)	Lymphocytosis with lymphadenopathy, with or without splenomegaly or hepatomegaly
High risk (stage III/IV)	Lymphocytosis with anemia or thrombocytopenia due to bone marrow involvement
Binet Staging System	
A	<3 areas of lymphadenopathy
B	≥3 areas of lymphadenopathy
C	Hemoglobin ≤10 g/dL and/or platelets <100,000/ μ L

Abbreviation: CLL, chronic lymphocytic leukemia.

TABLE 107-4 CLL International Prognostic Index

Risk Score		
VARIABLE	ADVERSE FACTOR	RISK SCORE
TP53 status	Deleted or mutated	4
IGHV mutational status	Unmutated	2
β_2 -microglobulin concentration	>3.5 mg/L	2
Clinical stage	Rai I–IV or Binet B–C	1
Age	>65 years	1
Implications of Risk Score		
RISK SCORE	RISK CLASSIFICATION	5-YEAR SURVIVAL (TRAINING SET DATA)
0–1	Low	93.2%
2–3	Intermediate	79.3%
4–6	High	63.3%
7–10	Very high	23.3%

Abbreviation: CLL, chronic lymphocytic leukemia.

presents with disease requiring urgent therapy, most times, these symptoms can be monitored over short periods to determine relatedness to CLL and need for therapy.

INITIAL THERAPY FOR CLL

Over the past decade, the initial therapy of CLL has dramatically changed. Whereas chemoimmunotherapy was once standard for all patients, now most patients are treated with oral therapies targeted against BTK or BCL2 with or without a CD20 monoclonal antibody. This continues to be an area of active investigation, with standards of care shifting rapidly. The major classes of these therapies are outlined here.

BTK Inhibitors BTK is an attractive target in CLL because, unlike other kinases in the BCR pathway, BTK does not have natural redundancy and is relatively selective for B cells, so inhibition leads to a predominant B cell-specific phenotype. The first-in-class BTK inhibitor is ibrutinib, which is relatively selective for BTK but also inhibits a number of structurally similar kinases. As initial therapy, ibrutinib was initially compared with chlorambucil (RESONATE 2 study), and there was an 84% lower risk of progression or death with ibrutinib, with 70% of ibrutinib-treated patients alive and progression-free at 5 years. Subsequent studies compared ibrutinib alone or with the anti-CD20 antibody rituximab to standard chemoimmunotherapy with fludarabine plus cyclophosphamide plus rituximab (FCR) in younger patients (<70 years; E1912 study) or bendamustine plus rituximab (BR) in older patients (≥65 years; A041202 study). In younger patients, ibrutinib plus rituximab (IR) showed increased progression-free survival (PFS) and overall survival (OS) when compared with FCR, with a 3-year PFS of 89% for IR compared with 71% for FCR. In older patients, ibrutinib alone as well as with rituximab showed superior PFS compared with BR, with 24-month PFS rates of 88% for IR, 87% for ibrutinib alone,

and 74% for BR. IR was not superior to ibrutinib alone, and OS was not different in this trial at 24 months. Side effects distinct to ibrutinib include arthralgias/myalgias, rash, diarrhea, dyspepsia, increased risk of bleeding (particularly when on anticoagulation therapy or with surgery), hypertension, and atrial fibrillation.

The second-generation BTK inhibitor acalabrutinib is more specific for BTK than ibrutinib and consequently shows better tolerability, with less incidence of atrial fibrillation, myalgias/arthralgias, and skin and nail changes than reported with ibrutinib. In the frontline setting, acalabrutinib and acalabrutinib plus obinutuzumab were compared with chlorambucil plus obinutuzumab. Both acalabrutinib alone and acalabrutinib with obinutuzumab showed superior 30-month PFS compared with chlorambucil plus obinutuzumab (82%, 90%, and 34%, respectively), with improved PFS for acalabrutinib plus obinutuzumab compared with acalabrutinib alone in an unplanned post hoc analysis.

BCL2 Inhibitor Venetoclax is an orally bioavailable, selective inhibitor of the antiapoptotic protein BCL2, which is upregulated in CLL. Unlike with BTK inhibitors, where many phase 3 studies support benefit over chemoimmunotherapy, only one study has been published with venetoclax. The CLL14 study compared venetoclax plus obinutuzumab (VO) to chlorambucil plus obinutuzumab in previously untreated patients with coexisting medical conditions. Unlike BTK inhibitors, which are administered continuously until disease progression, VO treatment is given for a 1-year fixed duration. At 3 years of follow-up, PFS was 82% in the VO group compared with 50% in the chlorambucil plus obinutuzumab group. No difference has been observed in OS with this follow-up. Side effects associated with venetoclax include tumor lysis syndrome, neutropenia, and nausea.

PI3K Inhibitors Inhibitors of PI3K delta have been studied in CLL due to the specificity of the delta isoform for B lymphocytes. Two agents, idelalisib and duvelisib, are approved for use in relapsed CLL, but trials of idelalisib in frontline CLL demonstrated toxicity that precluded further development in this area. Toxicities seen with idelalisib and duvelisib include pneumonitis, diarrhea/colitis, and transaminitis. More recently, a second-generation PI3K delta inhibitor umbralisib was combined with the anti-CD20 antibody ublituximab in the frontline setting and compared with chlorambucil plus obinutuzumab. Twenty-four-month PFS was 61% for ublituximab plus umbralisib compared with 40% with chlorambucil plus obinutuzumab. PI3K inhibitor-specific toxicities appear to be lower with umbralisib compared with idelalisib and duvelisib, but comparative trials are lacking. As outcome with this combination appears inferior to that with BTK inhibitors or BCL2 inhibitors, it is unlikely this treatment will be used in CLL outside of rare circumstances where other classes of drugs are contraindicated.

Chemoimmunotherapy For the most part, targeted therapy has supplanted chemoimmunotherapy in CLL. However, long-term follow-up of studies of FCR has demonstrated that a subset of patients treated with this regimen can have durable responses over 10 years, with a likely cure of CLL. This group is composed almost exclusively of patients with mutated *IGHV* and favorable cytogenetics. However, despite the efficacy of this regimen, short- and long-term toxicities limit its adaptability to many patients with *IGHV* mutated disease. Short-term toxicities are mostly related to myelosuppression and include neutropenia and infection. Long-term cytopenias are less common, but they do occur. Also, there is about a 3–5% risk of therapy-related myeloid neoplasm with this regimen that is almost always fatal. In the E1912 study of FCR versus IR, at follow-up, there was no difference in PFS or OS between FCR and IR for patients with mutated *IGHV*, suggesting that there may remain a place for this regimen in clinical practice. In addition, current studies are focused on limiting chemotherapy and/or adding novel agents in efforts to achieve cure but limit toxicity.

THERAPY OF RELAPSED CLL

Currently, the mainstays of treatment for relapsed CLL are the same classes as initial therapy. The optimal sequencing of targeted agents in

TABLE 107-5 Criteria for the Initiation of Therapy

Symptoms Indicating Need for Therapy in CLL

Evidence of progressive marrow failure (worsening of anemia or thrombocytopenia not due to autoimmune destruction)
Massive (≥6 cm below costal margin), progressive, or symptomatic splenomegaly
Massive (≥10 cm), progressive, or symptomatic lymphadenopathy
Progressive lymphocytosis with an increase of ≥50% over a 2-month period or lymphocyte doubling time <6 months
Autoimmune anemia or thrombocytopenia not responsive to standard therapy
Symptomatic or functional extranodal involvement
Constitutional symptoms (one or more of the following: unintentional weight loss ≥10% over 6 months, significant fatigue, fevers ≥100.5°F for 2+ weeks without infection, night sweats for >1 month without infection)

Abbreviation: CLL, chronic lymphocytic leukemia.

TABLE 107-6 Response Criteria in CLL

	LYMPHOCYTE COUNT	LYMPH NODES^a	SPLEEN/LIVER SIZE^b	BONE MARROW^c	PERIPHERAL BLOOD COUNTS
CR	<4000/ μ L	None >1.5 cm	Not palpable	Normocellular, <30% lymphocytes, no B lymphoid nodules	<ul style="list-style-type: none"> Platelet count >100,000/μL Hemoglobin >11 g/dL Neutrophils >1500/μL
PR	Decrease \geq 50% from baseline	Decrease \geq 50% from baseline	Decrease \geq 50% from baseline	Infiltrate \leq 50% of baseline	One of the following: <ul style="list-style-type: none"> Platelet count >100,000/μL or \geq50% from baseline Hemoglobin >11 g/dL or \geq50% from baseline Neutrophils >1500/μL or \geq50% from baseline
Stable disease	Not meeting CR/PR/PD criteria	Not meeting CR/PR/PD criteria	Not meeting CR/PR/PD criteria	Not meeting CR/PR/PD criteria	Not meeting CR/PR/PD criteria
PD	Increase \geq 50%	Increase \geq 50%	Increase \geq 50%		<ul style="list-style-type: none"> Platelet count \leq50% of baseline due to CLL Hemoglobin decrease >2 g/dL due to CLL

^aRefers to sum of the products of multiple lymph nodes evaluated by CT scan. ^bBased on physical examination. ^cBone marrow only required to confirm CR.

Abbreviations: CLL, chronic lymphocytic leukemia; CR, complete response; PD, progressive disease; PR, partial response.

CLL has not been established; however, the available data suggest that the sequence of either BTK inhibitor and then BCL2 inhibitor or the reverse are both acceptable. In a trial of venetoclax for patients who had relapsed after ibrutinib therapy, the overall response rate (ORR) was 65% with a median PFS of ~2 years in a very heavily pretreated patient population. Retrospective data of BTK inhibitor given after venetoclax suggests that this sequence is effective as well, with an ORR of 84% and median PFS of 32 months. PI3K inhibitors also have activity in relapsed CLL; however, activity following both BTK and BCL2 inhibitors is likely minimal. In addition, many new agents are in development in CLL including novel oral targeted therapies, antibodies, and immune-based treatments.

Immune Therapies Immune therapies in CLL are currently focused in the relapsed setting and include allogeneic stem cell transplantation, CAR-T therapy, and oral immunomodulatory agents such as lenalidomide.

Stem cell transplantation is a curative approach to CLL. Because most CLL patients are older and many have significant comorbidities, myeloablative transplants incur extensive morbidity and mortality, making them prohibitive in many individuals. Reduced-intensity conditioning (RIC) allogeneic transplants have been successfully incorporated into the treatment of patients up to ~75 years in age but still have a \geq 50% frequency of chronic graft-versus-host disease. This is still considered a standard treatment in CLL but has fallen out of favor with the introduction of well-tolerated novel agents, as well as clinical trials of CAR-T cells. CD19 CAR-T cell trials have not been as successful in CLL as they have been in other B cell malignancies due to the immunosuppression associated with the disease. Many current trials are focused on optimizing CD19 CAR-T cells by adding agents such as BTK inhibitors or PI3K inhibitors or modifying the CAR-T structure, and other studies are testing different targets outside of CD19. In addition, recent studies have shown that natural killer (NK) cell CAR cells also can induce clinical response in CLL patients. This area remains a focus of intense investigation in CLL.

ASSESSING RESPONSE TO THERAPY AND MINIMAL RESIDUAL DISEASE IN CLL

Following the completion of therapy or during therapy for indefinite targeted agents, response is initially assessed using physical examination and laboratory studies (Table 107-6). If residual disease is not detected using these methodologies, CT scans are used to assess response. Bone marrow biopsies with flow cytometry are indicated if no disease is detected to confirm complete response.

It has been established in various malignancies that complete tumor eradication is associated with longer survival. In CLL, if no malignant cells can be detected in the bone marrow down to a level of 1 CLL cell

in 10^4 leukocytes (0.01%), the patient is said to be negative for minimal residual disease (MRD). Following combination chemoimmunotherapy, eradication of MRD correlates with long-term survival and potentially cure in a subset of patients receiving FCR chemoimmunotherapy. Undetectable MRD in blood or bone marrow is also associated with improvement in PFS in venetoclax-based regimens. However, eradication of MRD has not been shown to be a meaningful endpoint with BTK or PI3K inhibitors as monotherapy. Higher sensitivity of 1 CLL in 10^6 leukocytes (0.0001%) can be obtained using next-generation sequencing methods such as ClonoSeq. This technique is currently available in clinical practice, although at this point, there are no data confirming that increased sensitivity is clinically meaningful, and studies are underway to support the need for this higher sensitivity with novel combination approaches of BTK/BCL2 inhibitor regimens.

CONCLUSION

CLL is treated only when it becomes symptomatic. At the time of therapy, FCR chemoimmunotherapy in a small subset of young patients with very-good-risk CLL is potentially curative. In the majority of patients with symptomatic CLL, targeted therapy directed at BTK or BCL2 can produce durable remissions and allow patients many years of disease-free survival.

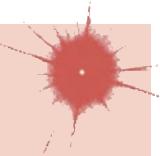
FURTHER READING

- BURGER JA: Treatment of chronic lymphocytic leukemia. *N Engl J Med* 383:460, 2020.
- FISCHER K et al: Venetoclax and obinutuzumab in patients with CLL and coexisting conditions. *N Engl J Med* 380:2225, 2019.
- HALLEK M et al: iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood* 131:2745, 2018.
- OAKES CC et al: DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet* 48:253, 2016.
- PUENTE XS et al: Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 475:101, 2011.
- SHANAFELT TD et al: Ibrutinib-rituximab or chemoimmunotherapy for chronic lymphocytic leukemia. *N Engl J Med* 38:432, 2019.
- SHARMAN JP et al: Acalabrutinib with or without obinutuzumab versus chlorambucil and obinutuzumab for treatment-naïve chronic lymphocytic leukemia (ELEVATE TN): A randomized, controlled, phase 3 trial. *Lancet* 395:1278, 2020.
- THOMPSON PA et al: Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival inIGHV-mutated chronic lymphocytic leukemia. *Blood* 127:303, 2016.
- WOYACH JA et al: Ibrutinib regimens versus chemoimmunotherapy in older patients with untreated CLL. *N Engl J Med* 380:1680, 2018.

108

Non-Hodgkin's Lymphoma

Caron A. Jacobson, Dan L. Longo



Non-Hodgkin's lymphomas (NHL) are cancers of mature B, T, and natural killer (NK) cells. They were distinguished from Hodgkin's lymphoma (HL) upon recognition of the Reed-Sternberg (RS) cell and differ from HL with respect to their biologic and clinical characteristics. Whereas ~80–85% of patients with HL will be cured of their lymphoma by chemotherapy with or without radiotherapy, the prognosis and natural history of NHL tends to be more variable. NHL can be classified as either a mature B-NHL or a mature T/NK-NHL depending on whether the cancerous lymphocyte is a B, T, or NK cell, respectively. Within each category are lymphomas that grow quickly and behave aggressively, as well as lymphomas that are more indolent, or slow growing, in nature. For a list of the World Health Organization (WHO) classification of lymphoid neoplasms, see **Table 108-1**.

■ EPIDEMIOLOGY AND ETIOLOGY

In 2020, >77,000 new cases of NHL were diagnosed in the United States, ~4% of all new cancers in both males and females, making it the seventh most common cause of cancer-related death in both women and men. The incidence is nearly 10 times the incidence of HL. There is a slight male-to-female predominance and a higher incidence for Caucasians than for African Americans. The incidence rises steadily with age, especially after age 40, but lymphomas are also among the most common malignancies in adolescent and young adult patients. The incidence of NHL has nearly doubled over the past 20–40 years and continues to rise by 1.5–2% each year. Patients with both primary

and secondary immunodeficiency states are predisposed to developing NHL. These include patients with HIV infection, patients who have undergone organ transplantation, and patients with inherited immune deficiencies and autoimmune conditions. The 5-year survival rates for NHL are 72% for Caucasians and 63% for African Americans.

The incidence of NHL and the patterns of expression of the various subtypes differ geographically and across age groups. T-cell lymphomas are more common in Asia than in Western countries, whereas certain subtypes of B-cell lymphomas such as follicular lymphoma (FL) are more common in Western countries. A specific subtype of NHL known as the angiocentric nasal T/NK cell lymphoma has a striking geographic occurrence, being most frequent in southern Asia and parts of Latin America. Another subtype of NHL associated with infection by human T-cell lymphotropic virus (HTLV) 1 is seen particularly in southern Japan and the Caribbean. Likewise, there are differences in the age-dependent incidence of NHL by histologic subtype, with aggressive lymphomas like diffuse large B-cell lymphoma (DLBCL) and Burkitt's lymphoma (BL) being the most common entities in children, and DLBCL and indolent lymphomas including FL being the most common forms in adults. The relative frequencies of the various types of lymphoid malignancies, including HL, plasma cell disorders, and lymphoid leukemias, is shown in **Fig. 108-1**.

A number of environmental factors have been implicated in the occurrence of NHL, including infectious agents, chemical exposures, and medical treatments. Several studies have demonstrated an association between exposure to agricultural chemicals and an increased incidence of NHL. Patients treated for HL can develop NHL; it is unclear whether this is a consequence of the HL or its treatment, especially radiation.

Several NHLs are associated with infectious agents (**Table 108-2**). Epstein-Barr virus (EBV) is associated with the development of BL in Central Africa and the occurrence of aggressive NHL in immunosuppressed patients in Western countries. The majority of primary central nervous system (CNS) lymphomas are associated with EBV. EBV infection is strongly associated with the occurrence of extranodal nasal NK/T-cell lymphomas in Asia and South America. HTLV-1 infects T cells and leads directly to the development of adult T-cell lymphoma (ATL) in a small percentage of patients infected as babies through ingestion of breast milk of infected mothers. The median age of patients with ATL is ~56 years; thus, HTLV-1 demonstrates a long latency from infection to oncogenesis (**Chap. 201**). Infection with HIV predisposes to the development of aggressive, B-cell NHL. This may be through overexpression of interleukin 6 by infected macrophages. Infection of the stomach by the bacterium *Helicobacter pylori* induces the development of gastric mucosa-associated lymphoid tissue (MALT) lymphomas. This association is supported by evidence that patients treated with antibiotics to eradicate *H. pylori* have regression of their MALT lymphoma. The bacterium does not transform lymphocytes to produce the lymphoma; instead, a vigorous immune response is made to the bacterium, and the chronic antigenic stimulation leads to the neoplasia. MALT lymphomas of the skin may be related to *Borrelia* sp. infections in Europe, those of the eyes to *Chlamydophila psittaci*, and those of the small intestine to *Campylobacter jejuni*. Chronic hepatitis C virus infection has been associated with the development of lymphoplasmacytic lymphoma and splenic marginal zone lymphoma (MZL). Human herpesvirus 8 is associated with primary effusion lymphoma in HIV-infected persons and multicentric Castleman's disease, a diffuse lymphadenopathy associated with systemic symptoms of fever, malaise, and weight loss.

In addition to infectious agents, a number of other diseases or exposures may predispose to developing lymphoma (**Table 108-3**). Diseases of inherited and acquired immunodeficiency as well as autoimmune diseases are associated with an increased incidence of lymphoma. The association between immunosuppression and induction of NHLs is compelling because if the immunosuppression can be reversed, a percentage of these lymphomas regress spontaneously. The incidence of NHL is nearly a hundredfold increased for patients undergoing organ transplantation necessitating chronic immunosuppression and is greatest in the first year posttransplant. About 30% of these arise as

TABLE 108-1 WHO Classification of Lymphoid Malignancies

B CELL	T CELL
Mature (peripheral) B-cell neoplasms	Mature (peripheral) T-cell neoplasms
Lymphoplasmacytic lymphoma (Waldenström's macroglobulinemia)	T-cell granular lymphocytic leukemia
Hairy cell leukemia	Adult T-cell leukemia/lymphoma (HTLV-1+)
Splenic marginal zone B-cell lymphoma	Extranodal NK/T-cell lymphoma, nasal type
Extranodal marginal zone B-cell lymphoma of MALT type	Enteropathy-associated T-cell lymphoma
Nodal marginal zone B-cell lymphoma	Hepatosplenic T-cell lymphoma
Follicular lymphoma	Subcutaneous panniculitis-like T-cell lymphoma
Mantle cell lymphoma	Mycosis fungoides
Diffuse large B-cell lymphoma (including subtypes)	Sezary syndrome
High-grade B-cell lymphoma with MYC and BCL2 and/or BCL6 rearrangements	Peripheral T-cell lymphoma NOS
High-grade B-cell lymphoma NOS	Angioimmunoblastic T-cell lymphoma
Burkitt's lymphoma/Burkitt's cell leukemia	Anaplastic large-cell lymphoma, ALK+
Primary mediastinal large B-cell lymphoma	Anaplastic large-cell lymphoma, ALK-
Plasmablastic lymphoma	
Primary effusion lymphoma	
HHV8+ DLBCL NOS	
Intravascular large B-cell lymphoma	
ALK+ large B-cell lymphoma	

Abbreviations: DLBCL, diffuse large B-cell lymphoma; HHV, human herpesvirus; HTLV, human T-cell lymphotropic virus; MALT, mucosa-associated lymphoid tissue; NK, natural killer; NOS, not otherwise specified; WHO, World Health Organization.

Source: Adapted from SH Swerdlow et al: *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, 5th ed. IARC, 2016.

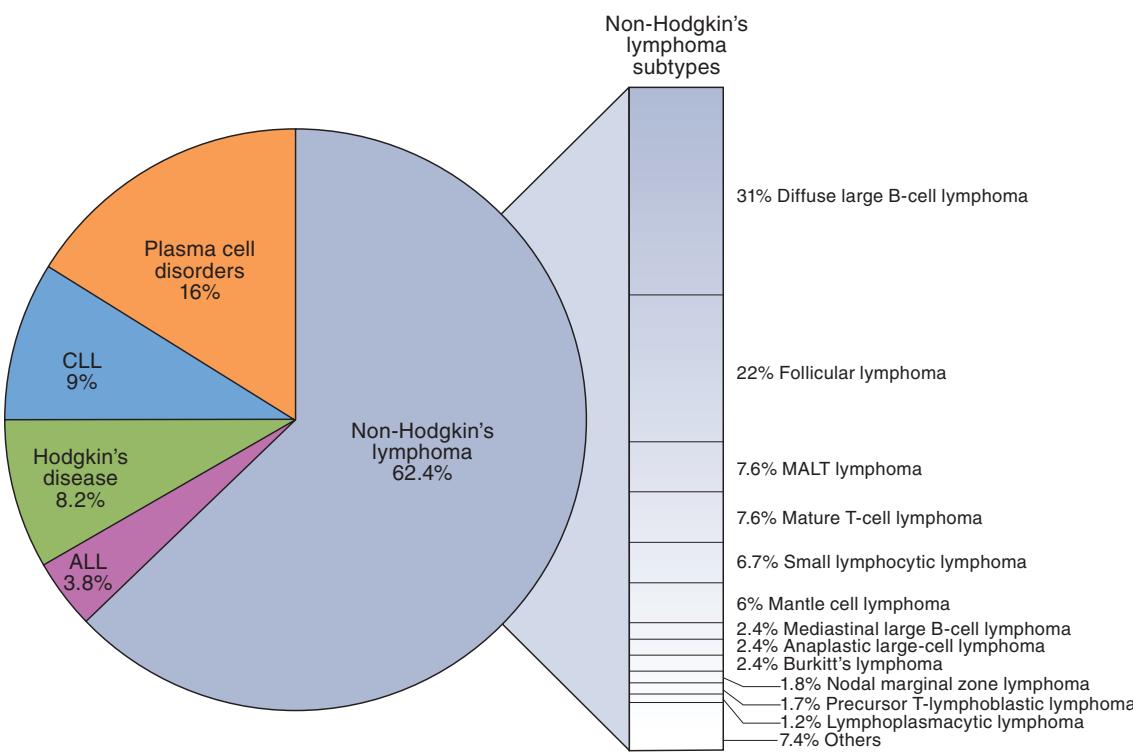


FIGURE 108-1 Relative frequency of lymphoid malignancies. ALL, acute lymphoid leukemia; CLL, chronic lymphocytic leukemia; MALT, mucosa-associated lymphoid tissue.

a polyclonal B-cell proliferation that evolves into a clonal B cell malignancy. The NHLs that occur in the context of immunosuppression or immunodeficiency, including HIV infection, are frequently associated with EBV. Histologically, DLBCLs are most frequently associated with immunosuppression and autoimmune diseases, although almost all histologies can be seen, especially MALT lymphomas in the context of autoimmune diseases such as Sjögren's syndrome and Hashimoto's thyroiditis. The rare inherited immunodeficiency diseases X-linked lymphoproliferative syndrome, Wiskott-Aldrich syndrome, Chédiak-Higashi syndrome, ataxia-telangiectasia, and common variable immunodeficiency syndrome are complicated by highly aggressive lymphomas. The elevated incidence of lymphoma in iatrogenic immunosuppression, AIDS, and autoimmune disease argues strongly for immune dysregulation contributing in the pathogenesis of some lymphomas. An increased risk of NHL has been observed in first-degree relatives with NHL, HL, or chronic lymphocytic leukemia (CLL). In large database studies, ~9% of patients with lymphoma or CLL have a first-degree relative with a lymphoproliferative disorder.

TABLE 108-2 Infectious Agents Associated with the Development of Lymphoid Malignancies

INFECTIOUS AGENT	LYMPHOID MALIGNANCY
Epstein-Barr virus	Burkitt's lymphoma Post-organ transplant lymphoma Primary CNS diffuse large B-cell lymphoma Hodgkin's lymphoma Extranodal NK/T-cell lymphoma, nasal type
HTLV-1	Adult T-cell leukemia/lymphoma
HIV	Diffuse large B-cell lymphoma Burkitt's lymphoma
Hepatitis C virus	Lymphoplasmacytic lymphoma
<i>Helicobacter pylori</i>	Gastric MALT lymphoma
Human herpesvirus 8	Primary effusion lymphoma Multicentric Castleman's disease

Abbreviations: CNS, central nervous system; HIV, human immunodeficiency virus; HTLV, human T-cell lymphotropic virus; MALT, mucosa-associated lymphoid tissue; NK, natural killer.

IMMUNOLOGY

All lymphoid cells are derived from a common hematopoietic progenitor that gives rise to lymphoid, myeloid, erythroid, monocyte, and megakaryocyte lineages. Through the ordered and sequential activation of a series of transcription factors, the cell first becomes committed to the lymphoid lineage and then gives rise to B and T cells.

About 90% of all lymphomas are of B cell origin. A cell becomes committed to B cell development when it expresses the master B lineage transcription factor PAX5, which ultimately results in a transcriptional program that leads to the rearrangement of its immunoglobulin genes, which involves chromosomal recombination as well as somatic hypermutation to create an immunoglobulin gene that is unique to that B cell. The sequence of cellular changes, including changes in cell-surface phenotype that characterizes normal B cell development, is shown in Fig. 108-2. Most B-cell lymphomas arise following the process of immunoglobulin gene recombination and somatic hypermutation, which leads to class switching and affinity maturation of the mature immunoglobulin, respectively, suggesting that it is the error-prone nature of these genetic events that contributes to oncogenesis. Certainly the frequency of chromosomal translocations that result in the activation of an oncogene or the inactivation of a tumor-suppressor gene in B-cell NHL may be the result of these normal cellular processes gone awry (see below). In addition, the key roles of the transcription factors MYC and BCL6 and the antiapoptotic protein BCL2 in the

TABLE 108-3 Diseases or Exposures Associated with Increased Risk of Development of Malignant Lymphoma

Inherited immunodeficiency disease	Autoimmune disease
Klinefelter's syndrome	Sjögren's syndrome
Chédiak-Higashi syndrome	Celiac sprue
Ataxia-telangiectasia syndrome	Rheumatoid arthritis and systemic lupus erythematosus
Wiskott-Aldrich syndrome	Chemical or drug exposures
Common variable immunodeficiency disease	Phenytoin
Acquired immunodeficiency diseases	Dioxin, phenoxy herbicides
Iatrogenic immunosuppression	Radiation
HIV-1 infection	Prior chemotherapy and radiation therapy
Acquired hypogammaglobulinemia	

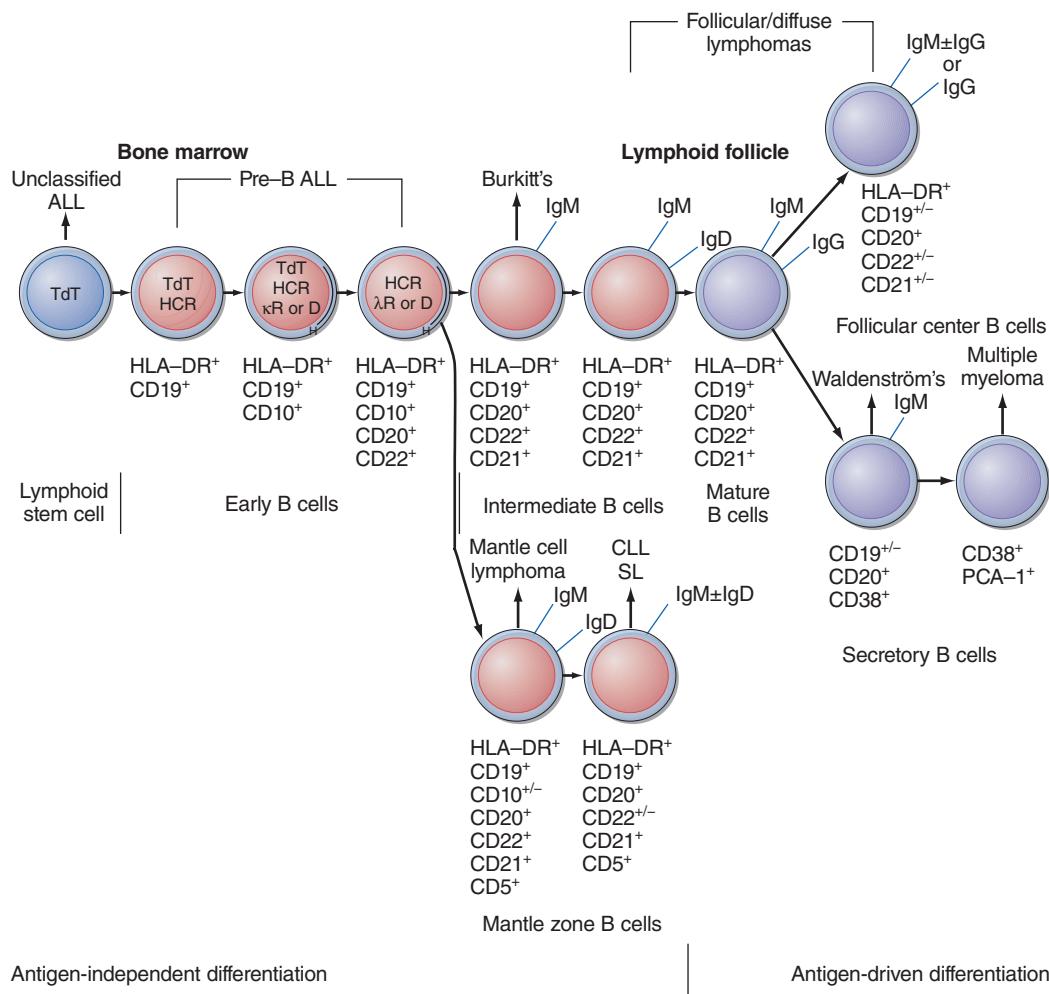


FIGURE 108-2 Pathway of normal B-cell differentiation and relationship to B-cell lymphomas. HLA-DR, CD10, CD19, CD20, CD21, CD22, CD5, and CD38 are cell markers used to distinguish stages of development. Terminal transferase (TdT) is a cellular enzyme. Immunoglobulin heavy chain gene rearrangement (HCR) and light chain gene rearrangement or deletion (κR or D, λR or D) occur early in B-cell development. The approximate normal stage of differentiation associated with particular lymphomas is shown. ALL, acute lymphoid leukemia; CLL, chronic lymphocytic leukemia; SL, small lymphocytic lymphoma.

process of B cell development explain why the genes encoding these proteins are commonly mutated in B-cell lymphomas.

A cell becomes committed to T-cell differentiation upon migration to the thymus and rearrangement of T-cell receptor (TCR) genes. This requires the expression of the T-cell master regulatory transcription factor, NOTCH-1. As in B cells, the development of the mature TCR involves the rearrangement and recombination of the TCR loci, which is error-prone and potentially oncogenic. The sequence of the events that characterize T-cell development is depicted in Fig. 108-3.

Although lymphoid malignancies often retain the cell-surface phenotype of lymphoid cells at particular stages of differentiation, this information is of little clinical or prognostic consequence. The so-called stage of differentiation of a malignant lymphoma does not predict its natural history. The antigen footprint, or immunophenotype, of the cell, however, is valuable diagnostically as it allows for the distinguishing of specific NHL subtypes. It can be detected by flow cytometry of single-cell suspension from blood, bone marrow, body fluid, or disaggregated tissue using fluorescently labeled antibodies against these antigens or by immunohistochemical staining of paraffin-embedded tissue sections with enzyme-linked antibodies against these antigens followed by a colorimetric reaction.

As already mentioned, malignancies of lymphoid cells are associated with recurring genetic abnormalities including chromosomal translocations and genetic mutations that may in part be the result of aberrant immunoglobulin or TCR development. While specific genetic abnormalities have not been identified for all subtypes of lymphoid malignancies, it is presumed that they exist. As previously discussed, B cells are even more susceptible to acquiring mutations during their

maturity in germinal centers; the generation of antibody of higher affinity requires the introduction of mutations into the variable region genes in the germinal centers. Given this, other nonimmunoglobulin genes, e.g., *bcl-6*, may acquire mutations as well. Likewise, many lymphomas contain balanced chromosomal translocations involving the antigen receptor genes; immunoglobulin genes on chromosomes 2, 14, and 22 in B cells; and T-cell antigen receptor genes on chromosomes 7 and 14 in T cells. The rearrangement of chromosome segments to generate mature antigen receptors must create a site of vulnerability to aberrant recombination. Examples of this type of event include the (8;14)(q24;q32) translocation in BL, involving the *MYC* proto-oncogene and the IgH gene; the (14;18)(q32;q32) translocation in FL, involving the *BCL2* proto-oncogene and the IgH gene; and the (11;14)(q13;q32) translocation in mantle cell lymphoma (MCL), involving the gene encoding cyclin D1 (*CCND1*) and the IgH gene. Less commonly, chromosomal translocations produce fusion genes that encode chimeric oncogenic proteins. Examples of this include the (2;5)(p23;q35) translocation involving the *ALK* and *NPM1* genes in anaplastic large-cell lymphoma (ALCL) and the t(11;18)(q21;q21) translocation involving the *API2* and *MLT* genes in MALT lymphoma. Table 108-4 presents the most common translocations and associated oncogenes for various subtypes of lymphoid malignancies.

Gene profiling using array technology allows the simultaneous assessment of the expression of thousands of genes. This technology provides the possibility to identify new genes with pathologic importance in lymphomas, the identification of patterns of gene expression with diagnostic and/or prognostic significance, and the identification of new therapeutic targets. Recognition of patterns of gene expression

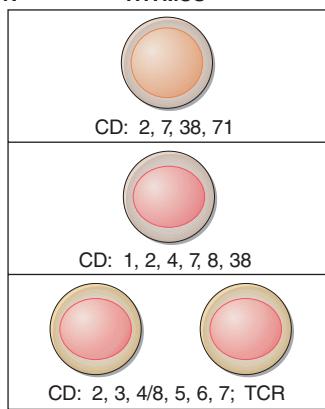
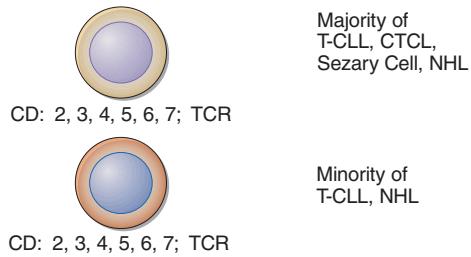
T-CELL DIFFERENTIATIONStage I
ProthymocyteStage II
ThymocyteStage III
Thymocyte**T-CELL MALIGNANCIES**Majority of
T-cell ALLMinority of T-ALL
Majority of T-LLMinority of T-LL
Rare T-ALL**PERIPHERAL BLOOD AND NODES**Mature T Helper
CellMature T Cytotoxic/
Suppressor CellMajority of
T-CLL, CTCL,
Sezary Cell, NHLMinority of
T-CLL, NHL

FIGURE 108-3 Pathway of normal T-cell differentiation and relationship to T-cell lymphomas. CD1, CD2, CD3, CD4, CD5, CD6, CD7, CD8, CD38, and CD71 are cell markers used to distinguish stages of development. T-cell antigen receptors (TCR) rearrange in the thymus, and mature T cells emigrate to nodes and peripheral blood. ALL, acute lymphoid leukemia; T-ALL, T-cell ALL; T-LL, T-cell lymphoblastic lymphoma; T-CLL, T-cell chronic lymphocytic leukemia; CTCL, cutaneous T-cell lymphoma; NHL, non-Hodgkin's lymphoma.

is complicated and requires sophisticated mathematical techniques. Early successes using this technology in lymphoma include the identification of previously unrecognized subtypes of DLBCL whose gene expression patterns resemble either those of follicular or germinal center B (GCB) cells or activated peripheral blood B cells (ABC). Patients whose lymphomas have a GCB-like pattern of gene expression have a considerably better prognosis than those whose lymphomas have a pattern resembling ABCs. This improved prognosis is independent of other known prognostic factors. These subcategories have been more specifically refined into five subcategories, using more advanced genetic sequencing techniques, that differ with respect to biology and driver genes, as well as prognosis, and may have important treatment implications in the future. Similar information is being generated in FL and MCL. The challenge remains to provide information from such techniques in a clinically useful time frame.

APPROACH TO THE PATIENT

Regardless of the type of lymphoid malignancy, the initial evaluation of the patient should include performance of a careful history and physical examination. These will help confirm the diagnosis, identify those manifestations of the disease that might require prompt attention, and aid in the selection of further studies to optimally characterize the patient's status to allow the best choice of therapy. It is difficult to overemphasize the importance of a carefully done history and physical examination. They might provide observations that lead to reconsidering the diagnosis, provide hints at etiology, clarify the stage, and allow the physician to establish rapport with the patient that will make it possible to develop and carry out a therapeutic plan.

The duration of symptoms and pace of symptomatic progression are important in distinguishing aggressive from more indolent lymphomas, as are the presence or absence of "B" symptoms, such as fevers, night sweats, or unexplained weight loss. Patients should be asked about localizing symptoms that may point toward lymphomatous involvement of specific sites, such as the chest, abdomen, or CNS. Comorbid diagnoses that may impact therapy or monitoring on therapy should be reviewed and acknowledged, including a history of diabetes or congestive heart failure. A physical examination should pay close attention to all the peripherally accessible sites of lymph nodes; the liver and spleen size; Waldeyer's ring; whether there is a pleural or pericardial effusion or abdominal ascites; whether there is an abdominal, testicular, or breast mass; and whether there is cutaneous involvement because all of these findings may influence further evaluation and disease management.

Laboratory studies should include a complete blood count, routine chemistries, liver function tests, and serum protein electrophoresis to document the presence of circulating monoclonal paraproteins. The serum β_2 -microglobulin level and serum lactate dehydrogenase (LDH) are important independent prognostic factors in NHL. Staging of certain diseases may involve a bone marrow biopsy; results of other laboratory and staging studies may also warrant a marrow evaluation. A lumbar puncture for evaluation of lymphomatous involvement may be indicated in the setting of concerning neurologic signs or symptoms or diseases that are high risk for CNS involvement. The latter may include disease involving the paranasal sinuses, testes, breast, kidney, adrenal glands, and epidural space, as well as highly aggressive histologies like BL. Since HIV and hepatitis B and C infection can be risk factors for developing NHL, and since treatment for some NHLs can result in the potentially life-threatening reactivation of hepatitis B, patients with a new diagnosis of NHL should be screened for these viruses as well.

Lymphoma histology and clinical presentation dictate which imaging studies should be ordered. Chest, abdominal, and pelvic computed tomography (CT) scans are essential for accurate staging to assess lymphadenopathy for indolent lymphomas, whereas positron emission tomography (PET) using ^{18}F -fluorodeoxyglucose

TABLE 108-4 Genetic Features of B- and T-Cell Lymphomas

GENETIC FEATURE	GENES	LYMPHOMA
t(8;14)	MYC/IgH	Burkitt's lymphoma
t(2;8)	MYC/Ig κ	
t(8;22)	MYC/Ig λ	
t(11;14)	BCL1 (CCND1)/IgH	Mantle cell lymphoma; multiple myeloma
t(14;18)	BCL2/IgH	Follicular lymphoma,
t(3;14)	BCL6/IgH	diffuse large B-cell lymphoma (DLBCL)
t(11;18)	API2/MALT1	MALT lymphoma
t(1;14)	BCL10/IgH	
t(14;18)	MALT1/IgH	
t(3;14)	FOXP1/IgH	
Trisomy 3	Unknown	Splenic marginal zone lymphoma
7q21 deletion	CDK6	
t(9;14)	PAX5/IgH	Lymphoplasmacytic lymphoma
6q21 deletion	Unknown	
inv(14)	TCR α /TCL1	Peripheral T-cell lymphoma, NOS; T-PLL
t(14;14)		
t(2;5)	NPM1/ALK	Anaplastic large-cell lymphoma (ALCL)
t(1;2)	TPM3/ALK	
t(2;3)	TFG/ALK	
t(2;17)	CTLC/ALK	
inv(2)	ATIC/ALK	
Trisomy 3	Unknown	Angioimmunoblastic T-cell lymphoma
Trisomy 5	Unknown	
Isochromosome 7q	Unknown	Hepatosplenitic T-cell lymphoma

Abbreviations: MALT, mucosa-associated lymphoid tissue; NOS, not otherwise specified; T-PLL, T-cell prolymphocytic leukemia.

TABLE 108-5 Staging Evaluation for Non-Hodgkin's Lymphoma

Physical examination
Documentation of B symptoms
Laboratory evaluation
Complete blood counts
Liver function tests
Uric acid
Calcium
Serum protein electrophoresis
Serum β_2 -microglobulin
Chest radiograph
CT scan of abdomen, pelvis, and usually chest
Bone marrow biopsy
Lumbar puncture in lymphoblastic, Burkitt's, and diffuse large B cell lymphoma with positive marrow biopsy
Gallium scan (SPECT) or PET scan in large-cell lymphoma

Abbreviations: CT, computed tomography; PET, positron emission tomography; SPECT, single-photon emission computed tomography.

(FDG-PET) is useful for aggressive lymphomas, including BL, DLBCL, plasmablastic lymphoma, and the aggressive T-cell NHLs. FDG-PET is highly sensitive for detecting both nodal and extranodal sites involved by NHL. The intensity of FDG avidity, or standardized uptake value (SUV), correlates with histologic aggressiveness, and may be useful in cases when disease transformation of an indolent lymphoma to a diffuse aggressive lymphoma is suspected. PET scanning can also differentiate between treated disease and active disease at the end of therapy in patients with residual masses on CT scans. Consensus recommendations regarding PET scanning were published as a result of an International Harmonization Project and state that PET should only be used for DLBCL and HL, that scanning during therapy should only be done as part of clinical trials, and that the end-of-treatment scan should not be done before 3 weeks but preferably 6–8 weeks after chemotherapy and 8–12 weeks after radiation or chemoradiotherapy. There is no evidence that long-term follow-up should include PET scanning. More recently, though, PET scan results at the end of therapy for FL have been associated with prognosis, with patients with residual PET-avid disease at the end of treatment having a poorer prognosis than those who are PET negative, and so it may be used for this prognostic purpose. Finally, magnetic resonance imaging (MRI) is useful in detecting bone, bone marrow, and CNS disease in the brain and spinal cord. The staging evaluation is outlined in **Table 108-5**.

The Ann Arbor staging system developed in 1971 for HL was adapted for staging NHLs (**Table 108-6**). This staging system focuses on the number of tumor sites (nodal and extranodal),

TABLE 108-7 International Prognostic Index for NHL**Five Clinical Risk Factors**

Age ≥60 years
Serum lactate dehydrogenase levels elevated
Performance status ≥2 (ECOG) or ≤70 (Karnofsky)
Ann Arbor stage III or IV
>1 site of extranodal involvement

For Diffuse Large B Cell Lymphoma

0, 1 factor = low risk	35% of cases; 5-year survival, 73%
2 factors = low-intermediate risk	27% of cases; 5-year survival, 51%
3 factors = high-intermediate risk	22% of cases; 5-year survival, 43%
4, 5 factors = high risk	16% of cases; 5-year survival, 26%

For Diffuse Large B Cell Lymphoma Treated With R-CHOP

0 factor = good	10% of cases; 4-year survival, 94%
1, 2 factors = intermediate	45% of cases; 4-year survival, 80%
3, 4, 5 factors = poor	45% of cases; 4-year survival, 53%

Abbreviations: ECOG, Eastern Cooperative Oncology Group; NHL, non-Hodgkin's lymphoma; R-CHOP, rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone.

location, and the presence or absence of systemic, or B, symptoms. Table 108-6 summarizes the essential features of the Ann Arbor system.

This anatomic based system is less useful in NHL, which disseminates widely, not in an ordered stepwise fashion. A majority of patients with NHL have advanced-stage disease at diagnosis. Apart from early-stage disease limited to a radiation field where local therapy with radiation is an option, all other disease is treated the same regardless of stage. Histology and clinical parameters at presentation are more important than stage with respect to prognosis. The International Prognostic Index (IPI) is perhaps the best predictor of outcome (**Table 108-7**). The IPI was developed based on the analysis of >2000 patients with aggressive NHLs treated with an anthracycline-containing regimen. Age (≤60 vs >60), serum LDH (≤ normal vs > normal), performance status (0 or 1 vs 2–4), stage (I or II vs III or IV), and extranodal involvement (<1 site vs >1 site) were identified as independently prognostic for overall survival (OS). A point is awarded for each risk factor and then summed, defining four risk groups: low (0 or 1), low-intermediate (2); high-intermediate (3); and high (4–5). The 5-year OS rates for patients with scores of 0 to 1, 2, 3, and 4–5 were 73, 51, 43, and 26%, respectively. The age-adjusted IPI separates patients ≤60 from patients >60. For the age-adjusted IPI, only stage, LDH, and performance status were important. Younger patients with 0, 1, 2, or 3 risk factors had 5-year survival rates of 83, 69, 46, and 32%, compared to 56, 44, 37, and 21% for older patients. When factoring in the introduction and clinical benefit of rituximab, the 4-year progression-free survival rates are 94, 80, and 53% for 0 and 1, 2, or 3 or more risk factors, respectively.

The Follicular Lymphoma International Prognostic Index (FLIPI) is a similar predictive model for FL, derived from the analysis of >4000 patients. Age >60, stage III/IV disease, the presence of >4 nodal sites, an elevated serum LDH concentration, and a hemoglobin <12 were identified as independent prognostic variables, and summation of each variable identified three risk groups. The median 10-year survival rates for patients with zero to one (low risk), two (intermediate risk), or three or more (high risk) of these adverse factors were 71, 51, and 36%, respectively. Similar disease-specific IPIs have been developed for MCL and peripheral T-cell lymphoma (PTCL) as well. These prognostic indices take into account the proliferative index and cell-surface markers, respectively.

Finally, as mentioned previously, gene expression profiling has identified DLBCLs with differential prognoses: GCB and ABC, where GCB-like DLBCL is associated with a significantly better

TABLE 108-6 Ann Arbor Staging for Lymphoma^a

STAGE	DESCRIPTION
I	Involvement of a single lymph node region (I) or single extranodal site (IE)
II	Involvement of two or more lymph node regions or lymphatic structures on the same side of the diaphragm alone (II) or with involvement of limited, contiguous, extralymphatic organ or tissue (IIE)
III	Involvement of lymph node regions on both sides of the diaphragm (III), which may include the spleen (IIIS), or limited, contiguous, extralymphatic organ or tissue (IIIE), or both (IIIES)
IV	Diffuse or disseminated foci of involvement of one or more extralymphatic organs or tissues, with or without associated lymphatic involvement

^aAll stages are further subdivided according to the absence (A) or presence (B) of systemic B symptoms including fevers, night sweats, and/or weight loss (>10% of body weight over 6 months prior to diagnosis).

OS. A more readily accessible immunohistochemical algorithm has been developed, based on the presence or absence of CD10, BCL6, and MUM1 that correlates closely with gene expression profiles and can differentiate the majority of GCB from non-GCB-like DLBCL. These profiles have prognostic importance but, to date, do not alter treatment recommendations for the primary treatment of DLBCL. Current clinical trials do stratify by DLBCL subtype, and it appears that agents like the Bruton tyrosine kinase (BTK) inhibitor ibrutinib and lenalidomide are most active in non-GCB DLBCL in the relapsed setting. Treatment may then be differentiated by these subtypes in the future.

CLINICAL FEATURES, TREATMENT, AND PROGNOSIS OF SPECIFIC NHL

MATURE B-CELL NEOPLASMS

B-cell NHLs can be characterized into two broad groups—those that behave aggressively, require immediate or urgent treatment with combination chemotherapy regimens, and are potentially curable; and those that are more indolent in nature, can be observed and treated only when they cause symptoms or signs of organ function impairment, are very responsive to therapy, but are not ultimately curable in the vast majority of cases. Among the aggressive diseases, the most common are NHL and DLBCL, and the most rapidly proliferative are NHL and BL. FL is the second most common NHL and the most common indolent NHL. Other indolent NHLs include MZL, lymphoplasmacytic lymphoma (LPL), and hairy cell leukemia (HCL). MCL is an intermediate-grade lymphoma that shares some characteristics with the aggressive lymphomas (fairly urgent need for treatment and aggressive upfront combination chemotherapy regimens), but like the indolent lymphomas, it is not readily curable with conventional-dose therapies.

Burkitt's Lymphoma Burkitt's lymphoma/leukemia (BL) is a rare disease in adults in the United States, making up <1% of NHL, but it makes up ~30% of childhood NHL. It is one of the fastest growing neoplasms, with a doubling time of <24 h. In general, it is a pediatric tumor that has three major clinical presentations. The endemic (African) form presents as a jaw or facial bone tumor that spreads to extranodal sites including ovary, testis, kidney, breast, and especially the bone marrow and meninges. The nonendemic form has an abdominal presentation with massive disease, ascites, and renal, testis, and/or ovarian involvement and, like the endemic form, also spreads to the bone marrow and CNS. Immunodeficiency-related cases more often involve lymph nodes and may present as acute leukemia. BL has a male predominance and is typically seen in patients <35 years of age.

On biopsy, there is a monotonous infiltration of medium-sized cells with round nuclei, multiple nucleoli, and basophilic cytoplasm with vacuoles. The proliferation rate is ~100%, and tingible body macrophages give rise to the classic “starry sky” appearance of this tumor ([Fig. 108-4](#)). Tumor cells are positive for B-cell antigens CD19 and CD20 and surface immunoglobulin. They are also uniformly positive for CD10 and BCL6 but negative for BCL2. Endemic BLs are EBV positive, whereas the majority of nonendemic BLs are EBV negative. BL is associated with a translocation involving MYC on chromosome 8q24 in >95% of the cases. The most common partners are chromosomes 14, 2, or 22, rearrangements that produce fusions of MYC with either the IgH (80%), kappa (15%), or lambda (5%) light chain genes, respectively.

While exquisitely chemosensitive, it is imperative that treatment for BL be initiated quickly given the rapid doubling time and high morbidity of this disease. There are several effective intensive combination chemotherapy regimens, all of which incorporate high doses of cyclophosphamide. Prophylactic therapy to the CNS is mandatory. Cure can be expected in 80–90% of patients when treated promptly and correctly. Dose-adjusted EPOCH-R (rituximab, infusional etoposide/vincristine/doxorubicin, cyclophosphamide, prednisone) is highly effective. Salvage therapy has been generally ineffective in patients whose disease progresses after upfront therapy, emphasizing the

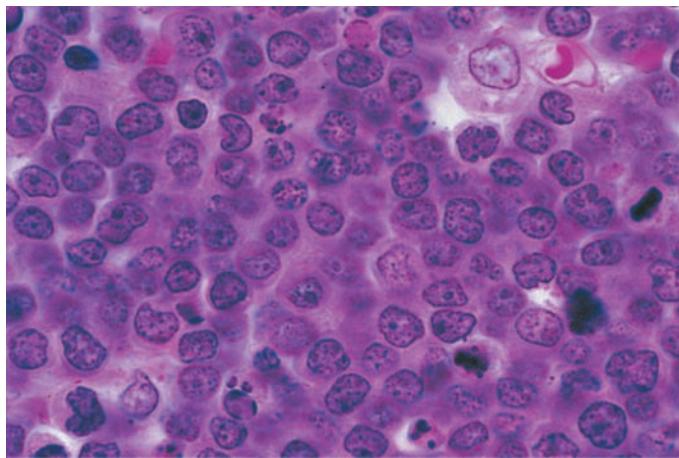


FIGURE 108-4 Burkitt's lymphoma. The neoplastic cells are homogeneous, medium-sized B cells with frequent mitotic figures, a morphologic correlate of high growth fraction. Reactive macrophages are scattered through the tumor, and their pale cytoplasm in a background of blue-staining tumor cells gives the tumor a so-called starry sky appearance.

importance of the initial treatment approach and referral to a tertiary cancer center with experience treating this disease.

Diffuse Large B-Cell Lymphoma

DLBCL is the most common histologic subtype of NHL diagnosed, representing about one-third of all cases. Previously felt to be “one disease,” it is now recognized as a heterogeneous collection of multiple entities. It is slightly more common in Caucasians and men, and the median age at diagnosis is 64. The relative risk (RR) of DLBCL is higher among people with affected first-degree relatives (RR 3.5-fold), and patients with congenital or acquired immunodeficiency, patients on immunosuppression, and patients with autoimmune disorders also have a higher risk of developing DLBCL, often EBV-related. The majority of patients present with advanced-stage disease, with only 30–40% of patients having stage I or II disease; ~40% of patients will have “B” symptoms, and 50% of patients will have an elevated LDH. Up to 40% of patients will have involvement of non-lymph node sites including bone marrow, CNS, gastrointestinal tract, thyroid, liver, and skin. Patients with extensive bone marrow involvement or involvement of the testes, breast, kidney, adrenal gland, paranasal sinus, or epidural space are at increased risk of CNS dissemination.

The tumor consists of a diffuse proliferation of large, atypical lymphocytes with a high proliferative index ([Fig. 108-5](#)). These cells typically express the B-cell antigens CD19, CD20, and CD79a. Expression of CD10 and BCL6 is consistent with the tumor cell being of germinal center origin (GCB), while the expression of MUM1 corresponds with

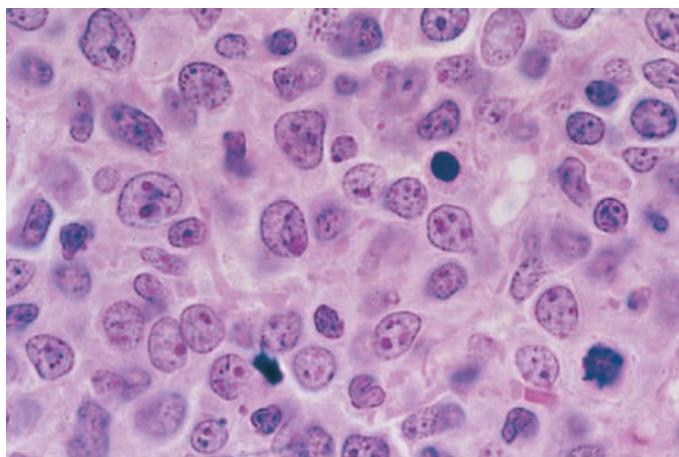


FIGURE 108-5 Diffuse large B-cell lymphoma. The neoplastic cells are heterogeneous but predominantly large cells with vesicular chromatin and prominent nucleoli.

the non-germinal center or activated B cell (ABC) subtype. BCL2 is overexpressed in anywhere from 25 to 80% of DLBCLs, whereas BCL6 is positive in more than two-thirds of cases, either as the result of translocations, gain of copy number, or promoter mutations. MYC is rearranged in 10% of DLBCLs, and ~20% of MYC-rearranged cases have concurrent BCL2 or BCL6 rearrangements, a combination referred to as “double-hit lymphoma.” These double-hit lymphomas are associated with an extremely poor prognosis with a median OS of only 12–18 months. Amplification and/or overexpression of MYC independent of rearrangements or amplification have also been described and are also associated with a poor, albeit better, prognosis.

Combination chemotherapy offers potentially curative therapy for DLBCL, regardless of the stage. The addition of the anti-CD20 antibody rituximab to cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) improved survival beyond CHOP alone and is the standard first-line chemotherapy for this disease. For patients with early-stage disease localized to a radiation field, treatment options include full-course chemotherapy with R-CHOP every 3 weeks for six cycles or abbreviated chemotherapy for three to four cycles followed by involved field radiotherapy. For advanced-stage DLBCL, therapy is with a full course of chemotherapy. On average, ~60–65% of patients with DLBCL can be expected to be cured with this approach, and the likelihood of cure is predicted by the IPI, gene expression profile cell of origin, and/or MYC cytogenetics and expression. Several studies have investigated alternative anthracycline-containing chemotherapy regimens and/or consolidation autologous stem cell transplantation in first remission for higher-risk disease without improvement over R-CHOP alone. Dose-adjusted R-EPOCH is one such regimen. Although this regimen did not appear to be better than R-CHOP for DLBCL in one multicenter clinical trial, it is often used to treat primary mediastinal large B-cell lymphoma and double-hit DLBCL based on results from phase 2 and retrospective studies, respectively. CNS prophylaxis with either intrathecal chemotherapy or high-dose systemic methotrexate and leucovorin rescue should be considered for patients with high risk of CNS dissemination. This includes patients with primary testicular involvement and breast involvement, as well as patients with several IPI risk factors and diffuse bone marrow involvement, renal involvement, or adrenal involvement. The use of CNS prophylaxis for disease involving the paranasal sinuses or the epidural space is less clear but may be considered.

Over one-third of patients will either have primary refractory disease or disease that relapses after first-line chemotherapy. These patients may still be cured with salvage chemotherapy regimens followed by autologous stem cell transplantation. However, patients with a poor performance status or advanced age who are not candidates for such an approach are often managed with palliative intentions. Radiation to symptomatic areas of disease can be transiently helpful. Less intensive chemotherapy with drugs like gemcitabine, cytarabine, or bendamustine can help control disease and symptoms for a limited period of time. These patients should be referred for clinical trials when applicable. For patients in whom more aggressive therapy is an option, treatment is with combination chemotherapy using various combinations of drugs primarily in order to identify patients with chemosensitive disease. Patients with chemosensitive disease have the greatest likelihood of benefiting from high-dose chemotherapy and autologous stem cell transplant, which improves response duration and survival over salvage chemotherapy alone and leads to long-term disease-free survival in ~40–50% of patients. For patients with chemorefractory disease, chimeric antigen receptor T cells (CAR-T cells) offer a potentially curative option. For this therapy, T cells are collected from a patient and are then genetically modified to express a receptor that will bind to a surface antigen expressed on the patient’s own tumor cells. In the case of B cell malignancies, CD19 has been targeted most commonly. After infusion, autologous CAR-T cells home to sites of disease and also persist over time. The CARs consist of an extracellular antigen recognition domain (typically a single chain Fv variable fragment from a monoclonal antibody) linked via a transmembrane domain to an intracellular signaling domain (usually the CD3 ζ endo-domain), resulting in the redirection of T cell specificity toward target

antigen-positive cells, and one or more costimulatory domains including CD28, 4-1BB, or OX40 to enhance cytokine secretion and effector cell expansion and prevent activation-induced apoptosis and immune suppression by tumor-related metabolites. Anti-CD19 CAR-T cells have been approved for the treatment of relapsed/refractory DLBCL following two prior systemic therapies. This would include patients with chemotherapy-insensitive disease following second-line salvage chemotherapy for whom autologous stem cell transplant is not an option or patients who relapse after autologous stem cell transplant. In this setting, the response rate of CAR-T cells is >80%, with >50% of patients achieving a complete response. These responses appear to be durable, with 40% of patients in remission at long-term follow-up.

Targeting CD19 with the monoclonal antibody tafasitamab in combination with lenalidomide also yielded high response rates and prolonged response durability, leading to approval of this regimen in relapsed disease. Reports of ongoing studies exploring bispecific antibodies that target CD20 on malignant B cells while also binding CD3 on T cells, thereby activating T cells to attack the malignant B cell, have been very promising in both aggressive and indolent B-cell NHL. The antibody-drug conjugate polatuzumab vedotin, which combines an anti-CD79b antibody with the microtubule toxin monomethyl aurostatin E (MMAE), was approved for the treatment of relapsed/refractory DLBCL in combination with bendamustine and rituximab based on the results of a randomized clinical trial against bendamustine and rituximab alone. The oral drug selinexor, a selective inhibitor of nuclear export, has modest activity in relapsed DLBCL as a single agent and is approved for this indication. These drugs, along with drugs such as lenalidomide alone or ibrutinib, should be viewed as a bridge to allogeneic stem cell transplant for eligible patients in whom curative therapy is the goal because they are unlikely to lead to durable or permanent remissions.

Other large B-cell lymphomas include intravascular large B-cell lymphoma, T-cell/histiocyte-rich large B-cell lymphoma, EBV-positive DLBCL of the elderly, and ALK-positive large B-cell lymphoma. Patients with the latter two diseases tend to have a poor prognosis, whereas the addition of rituximab to CHOP chemotherapy has dramatically improved outcomes with intravascular large B-cell lymphoma, and the outcomes in T-cell/histiocyte-rich large B-cell lymphoma are similar to DLBCL. R-CHOP remains the treatment of choice for each of these lymphomas.

Follicular Lymphoma FLs are the second leading NHL diagnosis in the United States and Europe and make up 22% of NHLs worldwide and at least 30% of NHLs diagnosed in the United States. This type of lymphoma can be diagnosed accurately on morphologic findings alone and has been the diagnosis in the majority of patients in therapeutic trials for “low-grade” lymphoma in the past.

Evaluation of an adequate biopsy by an expert hematopathologist is sufficient to make a diagnosis of FL. The tumor is composed of small cleaved and large cells in varying proportions organized in a follicular pattern of growth (**Fig. 108-6**). Confirmation of B-cell immunophenotype (monoclonal immunoglobulin light chain, CD19, CD20, CD10, and BCL6 positive, and CD5 and CD23 negative) and the existence of the t(14;18) and abnormal expression of BCL2 protein are confirmatory. While >85% of FLs will harbor a t(14;18) and overexpress the antiapoptotic protein BCL2, this genetic event is necessary but not sufficient for malignant transformation of the B lymphocytes, and multiple genetic events are required for the development of FL. Studies have identified the most common recurrent genetic events in FL, and they included mutations in several epigenetic modifying genes, including *MLL2*, *EZH2*, *CREBBP*, and *EP300*. The major differential diagnosis is between lymphoma and reactive follicular hyperplasia. The coexistence of DLBCL must be considered. Patients with FL are often subclassified, or graded, into those with predominantly small cells, those with a mixture of small and large cells, and those with predominantly large cells. The WHO classification adopted grading from I to III based on the number of centroblasts, or large cells, counted per high-power field (hpf): grade I, from 0 to 5 centroblasts/hpf; grade II, from 6 to 15 centroblasts/hpf; and grade III, >15 centroblasts/hpf. Grade III has

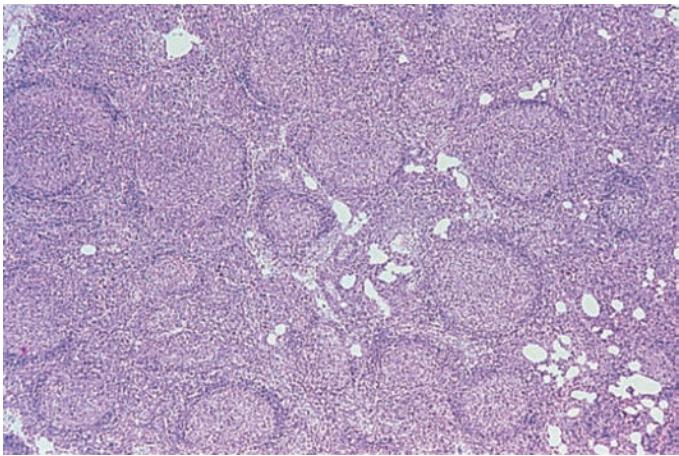


FIGURE 108-6 Follicular lymphoma. The normal nodal architecture is effaced by nodular expansions of tumor cells. Nodules vary in size and contain predominantly small lymphocytes with cleaved nuclei along with variable numbers of larger cells with vesicular chromatin and prominent nucleoli.

been subdivided into grade IIIa, in which centrocytes predominate, and grade IIIb, in which there are sheets of centroblasts. While this distinction cannot be made simply or very reproducibly, these subdivisions do have prognostic significance. Patients with FL with predominantly large cells have a higher proliferative fraction, progress more rapidly, and have a shorter OS with simple chemotherapy regimens. Grade IIIb FL is an aggressive disease and considered most similar to DLBCL and treated as such with curative intent.

The most common presentation for FL is with new, painless lymphadenopathy. Multiple sites of lymphoid involvement are typical, and unusual sites such as epitrochlear nodes are sometimes seen. However, essentially any organ can be involved, and extranodal presentations do occur. Most patients do not have an elevated LDH or fevers, night sweats, or weight loss, although histologic transformation to DLBCL does occur at a rate of ~3% per year and can be associated with these signs or symptoms. As discussed previously, prognosis is best predicted by the FLIPI. Staging is typically done with CT scans of the chest, abdomen, and pelvis, as well as the neck if neck disease is suspected, although PET/CT scans can be helpful in cases where disease transformation is suspected, as transformed disease will be more FDG avid than indolent disease, or for confirmation of early-stage disease, where definitive local therapy with radiation may be considered.

Although FL is highly sensitive to chemotherapy and radiotherapy, these therapies are usually not ultimately curative, except in the setting of early-stage disease. If the disease can be encompassed in a radiation field, involved field radiotherapy at a dose of 24–30 Gy may be curative, with 5-, 10-, and 15-year freedom from treatment failure rates of 72, 46, and 39%, and overall 5-, 10-, and 15-year survival rates of 93, 75, and 62%, respectively. If radiation therapy would not be tolerated or if a patient prefers not to receive radiation, observation is a reasonable alternative with a median time to treatment not reached at 7 years of follow-up in one study. Many of these patients are diagnosed incidentally or at a time when their lymphoma is not causing symptoms or signs of organ function impairment. Numerous studies have shown that treating patients with asymptomatic disease does not improve survival compared with a program of close observation, with treatment reserved for symptomatic disease progression or organ dysfunction. Thus, asymptomatic patients should be observed.

When treatment is indicated, there are a variety of treatment options, including the use of the monoclonal antibody against CD20, rituximab, alone or in combination with chemotherapy. Treatment decisions are often determined by the indication for treatment and/or by the volume of disease being treated. For patients requiring therapy for inflammatory or autoimmune phenomenon thought to be driven by FL, or for patients with low-volume disease, single-agent rituximab is associated with a response rate of ~70% and a median response

duration of >2 years. This response duration is improved with the addition of maintenance rituximab following a favorable response to rituximab induction therapy. For patients with a larger volume of disease at the time of treatment initiation, the addition of rituximab to chemotherapy regimens such as CHOP or cyclophosphamide, vincristine, and prednisone (CVP) has improved survival in this disease. The combination of bendamustine and rituximab (BR) has been compared to R-CHOP and results in longer response duration and less toxicity. Thus, BR has become the standard of care for the first-line therapy of medium- to high-volume FL. Similarly, the addition of maintenance rituximab following a good response to R-CHOP or R-CVP improves response duration when used in newly treated FL patients. A newer anti-CD20 antibody, obinutuzumab, has been tested in combination with chemotherapy in a randomized trial against rituximab plus chemotherapy in previously untreated FL. The obinutuzumab combinations resulted in improvements in minimal residual disease (MRD) negativity as well as progression-free survival at the expense of more infection and infusion reactions. Based on these results, both rituximab plus chemotherapy and obinutuzumab plus chemotherapy are options for untreated FL in need of treatment. The superiority of one over the other has not been established.

In patients with FL, the disease nearly always recurs following therapy, after which retreatment is again reserved for symptomatic disease or disease interfering with organ function. Single-agent rituximab or alternative chemotherapy regimens, with both rituximab and obinutuzumab, can again be employed. Both autologous and allogeneic hematopoietic stem cell transplantations yield high complete response rates in patients with relapsed FL, and long-term remissions can occur in 40 and 60% of patients, respectively. The latter is associated with considerable treatment-related morbidity and mortality and so is usually reserved for patients with multiply relapsed FL that is no longer responsive to chemotherapy. More targeted oral therapies like lenalidomide and the PI3 kinase inhibitors idelalisib, duvelisib, and copanlisib are active in both untreated and relapsed FL. Inhibitors of one of the most commonly mutated genes in FL, EZH2, have activity in both EZH2 mutated as well as unmutated lymphomas, and one, tazemetostat, is approved for this indication. Anti-CD19-directed CAR-T cell therapies are also being tested in FL, with complete responses seen in >80% of patients with multiply relapsed disease, and with many of those responses proving durable, albeit with limited follow-up. Longer follow-up is needed to determine if this may be a definitive treatment strategy for a subset of relapsed FL patients. On average, most patients will live with FL for 15–20 years, a number that is increasing given our improved understanding of the genetics and microenvironment of FL and the increasing number of drugs and therapies being tested in this disease. However, in addition to a high-risk FLIPI, patients who do not have a complete metabolic response by PET/CT scanning to their primary therapy and patients who relapse within 2 years of the completion of their primary chemotherapy tend to do poorly with chemotherapy.

Patients with FL have a high rate of histologic transformation to DLBCL (~3% per year). This is recognized ~40% of the time during the course of the illness by repeat biopsy and is present in almost all patients at autopsy. This transformation is usually heralded by rapid growth of lymph nodes—often localized—and the development of systemic symptoms such as fevers, sweats, and weight loss. When this happens in patients who have had previously untreated FL, treatment with R-CHOP chemotherapy, as for DLBCL, can be curative for the aggressive component while the FL may eventually recur. In patients with previously treated FL that transforms to DLBCL, prognosis is poor, and successful therapy with an aggressive combination chemotherapy regimen should be consolidated with an autologous stem cell transplant. Finally, as discussed previously, grade IIIb FL is more similar to DLBCL than it is to FL and should be treated as such.

Marginal Zone Lymphoma The second most common indolent B-cell NHL is MZL. There are three main types: splenic MZL, extranodal MZL of MALT, and nodal MZL.

Nodal MZL most closely resembles FL clinically, and much of the way we manage and treat it is based on studies done in FL. Tumor biopsies in this disease show parafollicular and perivascular infiltration by monocyteoid-appearance atypical lymphocytes with folded nuclear contours that are positive for CD19, CD20, and CD79a but negative for CD10 and largely negative for CD5. Some cases can have plasmacytoid differentiation and can be associated with a monoclonal expression of kappa or lambda light chains and with small monoclonal immunoglobulin spikes. Treatment is often similar to that of FL, with the exception that the BTK inhibitor ibrutinib is highly active in this disease, while largely disappointing in FL, and is a good treatment option for relapsed nodal MZL as well as other MZL subtypes.

Splenic MZL is largely a disease of older Caucasian patients; infection with hepatitis C is a risk factor for this disease, and treatment of hepatitis C can result in regression of the lymphoma. Patients present with a lymphocytosis with or without cytopenias and splenomegaly. Bone marrow involvement is common. Diagnosis can be made by flow cytometry of the peripheral blood; malignant lymphocytes will be positive for surface immunoglobulin, CD19, and CD20 and will generally lack CD5 and CD10. On peripheral smear, they have small nuclei and abundant cytoplasm with "shaggy" or villous projections. It can be differentiated from HCL by the absence of CD25, CD103, and annexin A1. Recurrent cytogenetic abnormalities include trisomy 3 and abnormalities of chromosome 7q. Therapy is indicated for symptomatic disease or significant cytopenias. Splenectomy is reasonable for selected patients with excellent relief of symptoms and cytopenias. Splenectomy is associated with an overall response rate of 85% and estimated progression-free survival and OS rates at 5 years of 58 and 77%, respectively. Single-agent rituximab can improve splenomegaly and cytopenias in >90% of patients. In a study of induction with weekly rituximab followed by maintenance, the response rate was 95%, with overall and progression-free survival rates at 5 years of 92 and 73%, respectively. Other options for therapy at relapse are similar to those used for FL and include retreatment with rituximab, alkylating agents, and purine analogues in combination with rituximab. The survival rate of patients is in excess of 70% at 10 years.

MALT lymphoma is an MZL lymphoma of extranodal tissue, most commonly the stomach, but other common sites include the skin, salivary glands, lung, small bowel, ocular adnexa, breasts, bladder, thyroid, dura, and synovium. It is associated with states of chronic inflammation due to either autoimmune diseases like Sjögren's syndrome or Hashimoto's thyroiditis or chronic infections with organisms like *H. pylori* (gastric), *Borrelia burgdorferi* (skin), *C. psittaci* (conjunctiva), *C. jejuni* (intestines), and hepatitis C virus. The essential pathologic feature of MALT lymphoma is the presence of lymphoepithelial lesions, which result from invasion of mucosal glands and crypts by the neoplastic lymphocytes. These cells are positive for CD19, CD20, and CD79a and negative for CD5 and CD10. Recurrent cytogenetic abnormalities include t(11;18), t(14;18), t(1;14), t(3;14), and trisomy 8. The t(11;18) is most common, occurring in up to 50% of MALT lymphomas. It results in the fusion of the apoptosis inhibitor 2 (*API2*) gene and the *MALT1* gene, resulting in activation of nuclear factor- κ B (NF- κ B). Unlike other indolent B-cell lymphomas, MALT lymphomas present most commonly with stage I or II disease. In these cases, radiation therapy may be curative. Alternatively, patients may respond to antibiotics for the associated underlying infection. Treatment of symptomatic or organ-impairing relapsed, refractory, or advanced-stage disease is similar to approaches used in FL with chemotherapy, immunotherapy, or chemoimmunotherapy.

Lymphoplasmacytic Lymphoma About 1% of all NHLs will be LPLs, which are indolent B-cell NHLs with lymphoplasmacytic differentiation, most commonly associated with a monoclonal IgM paraprotein. Nearly all patients will have stage IV disease at diagnosis with bone marrow involvement. Patients with high levels of circulating IgM paraproteins constitute a specific entity known as Waldenström's macroglobulinemia and can have symptoms due to hyperviscosity as a result of the circulating IgM. Activating mutations in MYD88, an

adaptor protein that is involved in signaling downstream of the Ig receptor leading to NF- κ B activation, are present in >90% of cases. Tumor biopsies are notable for proliferation of small lymphocytes, lymphoplasmacytic cells, and plasma cells, and malignant lymphocytes are positive for CD19, CD20, and surface IgM but generally negative for CD5 and CD10. Like the other indolent NHLs, treatment is indicated for disease that causes symptoms or interferes with organ function; hyperviscosity related to elevated serum IgM and paraneoplastic neuropathy are additional indications for therapy. Single-agent rituximab may be useful for low-volume disease but can be associated with a transient rise in serum IgM concentrations that can cause or exacerbate hyperviscosity. Chemoimmunotherapy with regimens such as BR and rituximab, cyclophosphamide, and dexamethasone is active, as are myeloma therapies such as bortezomib. Ibrutinib in combination with rituximab is highly active in this disease and is an option for both previously untreated and relapsed disease. Given that 85% of IgM remains intravascular, acute relief of hyperviscosity symptoms can be obtained by plasmapheresis. For recurrent disease, one can often use agents that were previously used. For patients with more refractory LPL, the mammalian target of rapamycin (mTOR) inhibitor everolimus and the oral BTK ibrutinib are active. Selected patients with relapsed disease are considered for high-dose therapy with autologous or allogeneic stem cell transplantation. The results seen are similar to those of other indolent lymphomas.

Mantle Cell Lymphoma MCL composes ~6% of NHLs. It is an intermediate-grade lymphoma that, like the indolent B-cell NHLs, is not curable with conventional therapies but, like the aggressive lymphomas, often requires more aggressive chemoimmunotherapy regimens with or without an autologous stem cell transplant to achieve a reasonable response duration. This therapy is not curative, however, and median survival with this disease is on the order of 5–10 years. An exception to this is a more indolent SOX11 variant that often presents with circulating disease with splenomegaly but without significant lymphadenopathy and with a low Ki67 (<10%). This subset behaves more like the indolent B-cell NHLs and can be observed until treatment is indicated by symptoms or organ function impairment. Similarly, there is a blastic variant with a high Ki67 index that is associated with a poor prognosis and a median OS of only 18 months. For other patients, prognosis is best predicted by the biologic MCL International Prognostic Index (MIPI), which factors in age, performance status, LDH, white blood cell count, and Ki67 expression to determine a risk group. This disease is more common in men, and the average age of diagnosis is 63. MCLs with a mutation in *TP53* or a complex karyotype are particularly high risk as well. Over two-thirds of patients will have stage IV disease, mostly with bone marrow and peripheral blood involvement, at the time of diagnosis. Another common extranodal site of involvement is the gastrointestinal tract, where diffuse lymphomatous polyposis may be seen.

The pathognomonic cytogenetic finding in MCL is t(11;14), which brings the gene for the cell cycle control protein cyclin D1 under the control of the immunoglobulin heavy chain gene promoter on chromosome 14. This translocation is present in >90% of cases. The remaining cases usually overexpress cyclin D2, cyclin D3, or cyclin E. Tumor cells also are positive for B cell markers CD19 and CD20, as well as CD5. They usually lack CD10 and CD23.

Therapies for MCL are evolving. Patients with localized disease might be treated with combination chemotherapy followed by radiotherapy; however, these patients are exceedingly rare. Similarly, patients with the indolent variant can be observed until disease progresses to cause symptoms or signs of organ function impairment. For the usual presentation with disseminated disease, standard lymphoma treatments like R-CHOP have been unsatisfactory, with the minority of patients achieving complete remission. The addition of high-dose cytarabine to an R-CHOP-like backbone with or without consolidation autologous stem cell transplantation in first remission has improved progression-free survival, but it has not elicited cures in this disease. These include the Nordic regimens and R-HyperCVAD (rituximab,

cyclophosphamide, vincristine, doxorubicin, dexamethasone, cytarabine, and methotrexate). BR has activity in this disease and is more effective and better tolerated than R-CHOP. Newer studies with short follow-up suggest that strategies that combine BR with cytarabine with or without autologous stem cell transplant may be effective and well tolerated. Maintenance rituximab, following a good response to induction chemotherapy or after autologous stem cell transplant, also improves outcomes over observation alone. For relapsed disease, the BTK inhibitors ibrutinib and acalabrutinib have single-agent activity with a response rate of almost 70% but a response duration of only 18 months. These drugs are being explored in combination with chemotherapy as well as with the BCL2 antagonist venetoclax. Anti-CD19-directed CAR-T cell therapies are approved for the treatment of relapsed/refractory MCL; two-thirds of patients who had progressed after chemoimmunotherapy (with or without an autologous stem cell transplant) and BTK inhibition have achieved complete responses, many of which are durable through limited follow-up. As in FL, longer follow-up is needed to determine if some of these patients may be cured, which would make this the only curative therapy for this disease outside of an allogeneic stem cell transplantation. Drugs such as lenalidomide, venetoclax, bortezomib, and temsirolimus can similarly induce transient partial responses. Appropriate patients who respond to salvage therapy, with the exception of CAR-T cell therapy, should be considered for allogeneic stem cell transplant, which can lead to long-term disease-free survival in 30–50% of patients.

MATURE (PERIPHERAL) T CELL DISORDERS

Mature T cell disorders include cutaneous lymphomas, such as mycosis fungoïdes, and the PTCLs, some of which are distinguished based on specific clinical presentations or contexts or by molecular or biologic features, but many of which fall into the category of PTCL not otherwise specified (NOS). T-cell NHLs are significantly rarer than B-cell NHLs, and as such, our understanding of their biology is less advanced and our therapies are less well developed. While some T-cell lymphomas, like mycosis fungoïdes, can behave indolently and some, like ALK-positive ALCL, can be cured with chemotherapy, the majority are associated with a poor prognosis. The advent of genomic technologies is enhancing our ability to understand the genetic and biologic basis of these neoplasms.

Mycosis Fungoïdes Mycosis fungoïdes is also known as cutaneous T-cell lymphoma. This lymphoma is more often seen by dermatologists than internists. The median age of onset is in the mid-fifties, and the disease is more common in males and in blacks.

Mycosis fungoïdes is an indolent lymphoma, with patients often having several years of eczematous or dermatitic skin lesions before the diagnosis is finally established. The skin lesions progress from patch stage to plaque stage to cutaneous tumors. Early in the disease, biopsies are often difficult to interpret, and the diagnosis may only become apparent by observing the patient over time. Adenopathy may reflect involvement with mycosis fungoïdes or be read as dermatopathic change. In advanced stages, the lymphoma can spread to lymph nodes and visceral organs. Patients with this lymphoma may develop generalized erythroderma and circulating tumor cells, called *Sézary's syndrome*.

Rare patients with localized early-stage mycosis fungoïdes can be cured with radiotherapy, often total-skin electron beam irradiation. More advanced disease has been treated with topical glucocorticoids, topical nitrogen mustard, phototherapy, psoralen with ultraviolet A (PUVA), extracorporeal photopheresis, retinoids (bexarotene), electron beam radiation, interferon, antibodies, fusion toxins, histone deacetylase inhibitors, brentuximab (for CD30+ disease), and systemic cytotoxic therapy. Mogamulizumab, an anti-CCR4 antibody, has activity in this disease and has been approved by the U.S. Food and Drug Administration for this indication. Unfortunately, these treatments are palliative.

Peripheral T-Cell Lymphoma, Not Otherwise Specified PTCLs include a number of entities, which constitute 15% of all NHLs in adults. PTCL NOS, which composes 6% of all NHLs, is the

term used for cases that are not other entities defined in the WHO classification. Named varieties include ALCL, angioimmunoblastic T-cell lymphoma (AITL), hepatosplenic T-cell lymphoma, enteropathy-associated T-cell lymphoma, and subcutaneous panniculitis T-cell lymphoma. PTCL NOS is a disease of older individuals, with a median age at presentation of 65, and the majority of patients will have advanced-stage disease at diagnosis, with involvement of the bone marrow, liver, spleen, and skin being common. Associated “B” symptoms and pruritis are also common. These lymphomas can be associated with a reactive eosinophilia as well as hemophagocytic syndrome. The IPI has been applied to PTCL NOS and provides some assessment of outcomes, but even the low-risk group has a median OS of just >2 years.

This diagnostic category is a collection of heterogeneous lymphomas that vary widely and lack typical findings of other specific PTCL subgroups. Because of this heterogeneity, histology, immunophenotype, and genetics are variable. Often lymph nodes are effaced by atypical lymphoid cells of various sizes, sometimes associated with vascular proliferation or an infiltrate of eosinophils and/or macrophages. As most of these lymphomas behave aggressively, note is often made of mitotic and apoptotic figures as well as geographic necrosis. The cells often are positive for CD3, and the majority of PTCL NOS is positive for CD4 rather than CD8, but some are negative for both markers. There can be loss of more mature T-cell markers like CD5 and CD7, and this is associated with a more aggressive course. There are some recurrent translocations, including t(7;14), t(11;14), inv(14), and t(14;14), all of which involve the TCR genes.

The most common primary therapy for PTCL NOS involves a CHOP-like chemotherapy backbone—either CHOP alone or CHOP in combination with etoposide (CHOEP). The latter may provide the most benefit to younger patients and patients with more favorable disease risk factors. Brentuximab in combination with cyclophosphamide, doxorubicin, and prednisone (CHP) has been tested in a randomized clinical trial against CHOP in CD30+ T-cell lymphomas; progression-free survival was improved with the brentuximab-containing arm, and this was most pronounced for patients with ALCL (see below). Autologous stem cell transplant has been investigated for patients in their first remission and does seem to improve progression-free survival in certain contexts. Drugs such as gemcitabine, bendamustine, and pralatrexate have activity in relapsed disease, as do the histone deacetylase inhibitors romidepsin and belinostat. The PI3 kinase inhibitor duvelisib is being investigated in these diseases with early signals of activity. All of these agents are associated with transient responses in a minority of patients. Patients should be considered for clinical trials. For patients who do achieve remission, reduced-intensity allogeneic stem cell transplantation can yield long-term nonrelapse survival rates of ~40–50%.

Angioimmunoblastic T-Cell Lymphoma AITL constitutes ~20% of T-cell NHLs and ~4% of all NHLs diagnosed. Patients present with a variety of signs and symptoms, most often including lymphadenopathy, hepatosplenomegaly, “B” symptoms, rash, polyarthritis, and hemolytic anemia. Over 80% of patients have advanced-stage disease at diagnosis, and bone marrow involvement is common. Polyclonal hypergammaglobulinemia is common, as are elevated LDH, eosinophilia, a positive Coombs test, and opportunistic infections.

On biopsy, lymph nodes are effaced by a polymorphous infiltrate of lymphocytes, ranging in size and shape, and of immunoblasts. The neoplastic lymphocytes are positive for CD3 as well as CXCL13, PD-1, CD10, and BCL6, most closely resembling CD4-positive follicular helper T cells. There is an expanded follicular dendritic cell network surrounding tumor cells. Scattered immunoblasts are often EBV positive and may give rise to secondary EBV-positive B-cell lymphomas at a later time. Genetic analysis of this disease has revealed recurrent mutations in *TET2* (76%), *DNMT3* (33%), and *IDH2* (20%).

There is a subset of AITL that can remit with immunosuppression with agents like glucocorticoids or methotrexate. Most patients, however, will need combination chemotherapy with regimens like those used in PTCL NOS. Median response duration is short, and median OS

is only 15–36 months. Treatment of relapsed disease is similar to that of relapsed PTCL NOS.

Anaplastic Large-Cell Lymphoma ALCL is the next most common T-cell lymphoma after AITL but is more common in children, accounting for up to 10% of pediatric lymphomas. Approximately 40–60% of cases harbor t(2;5), which fuses a portion of the nucleolar protein nucleophosmin-1 (*NPM1*) gene to a part of the anaplastic lymphoma kinase (*ALK*) gene, the product of which has constitutive tyrosine kinase activity. These patients have a much more favorable prognosis compared to ALK-negative ALCL, akin to that of DLBCL. There is an additional, more indolent and favorable subtype that occurs in the breast tissue of patients with breast implants, and there is a cutaneous variant. In general, this is a disease that is more common in men. ALK-positive disease is a disease of younger patients, with a median age at diagnosis of 34 years, whereas the median age at diagnosis of ALK-negative patients is 58. With the exception of the cutaneous variant and the variant associated with breast implants, most patients present with rapidly growing lymphadenopathy with or without extranodal involvement; “B” symptoms are common.

Most cases of ALCL involve large atypical lymphocytes with horsehoe-shaped nuclei with prominent nucleoli (“hallmark” cells). Tumor cells tend to be localized within the lymph node sinuses, and almost all are positive for CD30 but negative for CD15. A majority will also express CD3, CD25, CD43, and CD4. ALK-rearranged ALCL can be diagnosed by fluorescence in situ hybridization (FISH) cytogenetics for t(2;5) or by immunohistochemical staining for ALK.

ALCL is generally treated with CHOP, although like PTCL NOS, CHOEP may benefit younger patients, particularly with ALK-positive disease. Overall, ALCL has a better prognosis than PTCL, and this is particularly true for ALK-positive disease, which has an 8-year OS rate of 82%, versus 49% for ALK-negative disease. Relapsed ALK-positive ALCL is treated similarly to relapsed DLBCL, with salvage combination chemotherapy to identify chemotherapy sensitivity followed by autologous stem cell transplant. For patients with chemotherapy-insensitive disease or for ALK-negative disease, the conjugated anti-CD30 antibody to MMAE brentuximab is highly active, with a response rate of 86% and a complete response rate of 57%. As mentioned earlier, brentuximab in combination with CHP chemotherapy is an approved frontline regimen for the treatment of CD30+ T-cell lymphomas, including ALCL. The ALK inhibitors, including crizotinib, are active in refractory ALK-positive ALCL with excellent outcomes.

Other PTCL Subtypes Enteropathy-associated T-cell lymphoma, hepatosplenic T-cell lymphoma, and subcutaneous panniculitis-like T-cell lymphoma are other less common PTCL subtypes. *Enteropathy-type intestinal T-cell lymphoma* is a rare disorder. Type I occurs in patients with a history of gluten-sensitive enteropathy and is associated with HLADQA1*0501, DQB1*0201; a gluten-free diet can prevent the development of this lymphoma. Type II is not associated with celiac disease and may be a separate disease entity. Patients are frequently cachectic and sometimes present with intestinal perforation. The prognosis is poor, with a median survival of 10 months. Therapy is often with combination chemotherapy, including high-dose methotrexate, and autologous stem cell transplant in first remission.

Hepatosplenic γδ T-cell lymphoma is a systemic illness that presents with sinusoidal infiltration of the liver, spleen, and bone marrow by malignant T cells. Tumor masses generally do not occur. The disease is associated with systemic symptoms and is often difficult to diagnose. Recurrent genetic events include isochromosome 7q and trisomy 8. Treatment outcome is poor, but regimens that include ifosfamide, such as ifosfamide, carboplatin, and etoposide (ICE) or ifosfamide, etoposide, and cytarabine (IVAC), are associated with better outcomes in small series of patients. Responding patients should be considered for allogeneic stem cell transplantation.

Subcutaneous panniculitis-like T-cell lymphoma is a rare disorder that is often confused with panniculitis. Patients present with multiple subcutaneous nodules, which progress and can ulcerate.

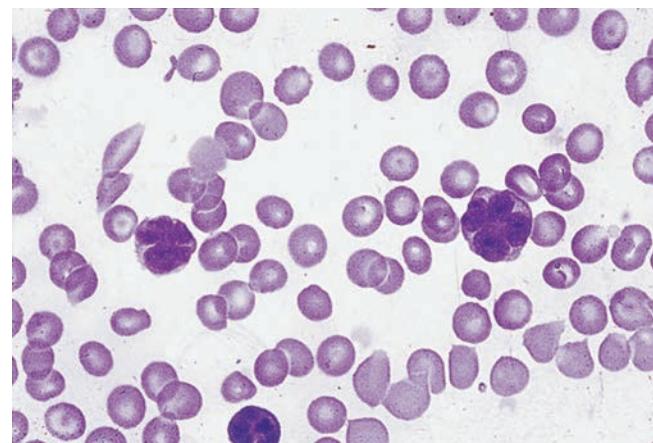


FIGURE 108-7 Adult T-cell leukemia/lymphoma. Peripheral blood smear showing leukemia cells with typical “flower-shaped” nucleus.

There is a more indolent form that tends to express α/β TCRs and can be managed with immune suppression, whereas lymphomas that express γ/δ TCRs are more aggressive and are associated with a worse prognosis and coincident hemophagocytic syndrome. This is a disease of young men in their fifth and sixth decades of life. Patients with aggressive disease are managed with multiagent chemotherapy, and responding patients should be considered for allogeneic stem cell transplantation.

Adult T-Cell Leukemia/Lymphoma Adult T-cell leukemia/lymphoma (ATLL) is a disease that is most prevalent in Japan and the Caribbean basin. It is a neoplasm that is driven by HTLV-1, often contracted through the breast milk of infected mothers. The average age at diagnosis is 60, so there is a long latency between viral infection and viral transformation, and only 4% of infected patients will develop the disease. This suggests that HTLV-1 may not be sufficient to cause the malignant phenotype. There are four disease variants: acute (60% of patients), lymphomatous (20% of patients), chronic (15% of patients), and smoldering (5% of patients); prognosis varies across these groups, with median survival times of 6, 10, and 24 months, and not yet reached, respectively. Presentation depends on the subtype, but most commonly, patients present with circulating disease and bone marrow involvement, hypercalcemia, lytic bone lesions, lymphadenopathy, hepatosplenomegaly, skin lesions, and opportunistic infections.

The pathognomonic finding is the malignant “flower cell” that is positive for CD4 and CD25, as well as CD2, CD3, and CD5 but lacking CD7 (Fig. 108-7). Combination chemotherapy is generally used, but for patients fortunate enough to respond, response durations are very short. Other active agents in this disease include the antiretroviral agent zidovudine, interferon α, and arsenic. In any patients who do respond to therapy, allogeneic stem cell transplant should be considered.

Extranodal NK/T-Cell Lymphoma, Nasal Type Extranodal NK/T-cell lymphoma, nasal type, is a lymphoma that is associated with EBV infection in nearly all cases and more common in Asia and native populations in Peru. It usually presents with a mass and obstructive symptoms in the upper aerodigestive tract with occasional extranodal sites, but over two-thirds of patients will have localized disease. It is more common in men, and the median age at diagnosis is 60. This disease has its own prognostic score, which takes into account the presence or absence of “B” symptoms, disease stage, whether LDH is elevated, and whether there is lymph node involvement. EBV viral load at diagnosis and at the end of therapy is also predictive.

Treatment for early-stage disease is usually with combined-modality therapy of chemotherapy (commonly using etoposide, ifosfamide, cisplatin, and dexamethasone) and intensity-modulated radiation therapy (50–55 Gy), and patients with localized disease involving the nasal passages do quite well, with 3-year OS of ~85%. Patients with

more advanced-stage disease do poorly, with disseminated extranodal relapse occurring frequently, and the median OS is only 4.3 months. The most commonly used treatment regimen is the SMILE regimen (dexamethasone, methotrexate, ifosfamide, L-asparaginase, and etoposide).

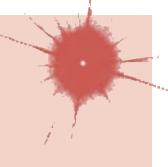
FURTHER READING

- HANEL W, EPPERLA N: Evolving therapeutic landscape in follicular lymphoma: a look at emerging and investigational therapies. *J Hematol Oncol* 14:104, 2021.
- ROSCHEWSKI M et al: Multicenter study of risk-adapted therapy with dose-adjusted EPOCH-R in adults with untreated Burkitt lymphoma. *J Clin Oncol* 38:2519, 2020.
- SAKATA-YANAGIMOTO M et al: Molecular understanding of peripheral T-cell lymphomas, not otherwise specified (PTCL, NOS): A complex disease category. *J Clin Exp Hematop* 61:61, 2021.
- SILKENSTEDT E, DREYLING M: Mantle cell lymphoma—advances in molecular biology, prognostication, and treatment approaches. *Hematol Oncol* 39 Suppl 1:31, 2021.

109

Hodgkin's Lymphoma

Caron A. Jacobson, Dan L. Longo



Hodgkin's lymphoma (HL) is a malignancy of mature B lymphocytes. It represents ~10% of all lymphomas diagnosed each year. The majority of HL diagnoses are classical HL (cHL), but there is a second subtype of HL, nodular lymphocyte-predominant HL (NLPHL). While this diagnosis does resemble cHL morphologically in certain respects, there is some evidence that it is more related to the indolent B-cell non-Hodgkin's lymphomas (NHLs) biologically than it is to cHL. The majority of this chapter will be specific to cHL, with a discussion of NLPHL at the end.

cHL is one of the success stories of modern oncology. Until the advent of extended-field radiotherapy in the mid-twentieth century, it was a highly fatal disease of young people. Radiation therapy cured some patients with early-stage disease, and the introduction of multiagent chemotherapy in the 1970s resulted in further improved cure rates, both for patients with early- and advanced-stage disease. Cure rates now are >85%. The new challenge in the treatment of HL is late therapy-related toxicity, including a high rate of secondary malignancies and cardiovascular disease. Current clinical trials are aimed at minimizing this risk while preserving efficacy.

EPIDEMIOLOGY AND ETIOLOGY

HL is of B-cell origin. The incidence of HL appears fairly stable, with 8480 new cases diagnosed in 2020 in the United States. HL is more common in whites than in blacks and more common in males than in females. A bimodal distribution of age at diagnosis has been observed, with one peak incidence occurring in patients in their twenties and the other in those in their eighties. Some of the late age peak may be attributed to confusion among entities with similar appearance such as anaplastic large-cell lymphoma and T-cell/histiocyte-rich B-cell lymphoma. There are four distinct subtypes of cHL that are differentiated based on their histopathologic features (Table 109-1): nodular sclerosis, mixed cellularity, lymphocyte-rich, and lymphocyte-depleted. Patients in the younger age groups diagnosed in the United States largely have the nodular sclerosing subtype of HL. Elderly patients, patients infected with HIV, and patients in developing countries more commonly have mixed-cellularity HL or lymphocyte-depleted HL. Together, nodular sclerosis and mixed-cellularity types account for nearly 95% of cases. Infection by HIV is a risk factor for developing

TABLE 109-1 World Health Organization Classification of Hodgkin's Lymphoma

Nodular lymphocyte-predominant Hodgkin's lymphoma
Classical Hodgkin's lymphoma
Nodular sclerosis
Lymphocyte-rich
Mixed cellularity
Lymphocyte-depleted

HL. In addition, an association between infection by Epstein-Barr virus (EBV) and HL has been suggested. A monoclonal or oligoclonal proliferation of EBV-infected cells in 20–40% of the patients with HL has led to proposals for this virus having an etiologic role in HL. However, the matter is not settled definitively. Viral oncogenesis appears to play a greater role in HIV-related cHL: EBV can be detected in nearly all cases of HIV-associated cHL, compared to only one-third of cases of non-HIV-associated cHL. Reed-Sternberg (HRS) cells are the malignant cells in HL. HRS cells in HIV-associated cHL express the EBV-transforming protein latent membrane protein 1 (LMP-1), and the EBV genomes from multiple disease sites in the same HIV-associated cHL patient are episomal and clonal, suggesting that EBV is directly involved in early lymphomagenesis.

Histologically, the HRS cell is diagnostic of cHL (Fig. 109-1). These cells are large cells with abundant cytoplasm with bilobed and/or multiple nuclei. By immunohistochemistry, they are often PAX-5 positive but have low to no expression of other B-cell antigens like CD19 and CD20. They express CD15 and CD30 in 85 and 100% of cases, respectively. These cells, though, comprise <1% of the tumor cellularity, with the majority of the tumor made up of a surrounding infiltrate of polyclonal lymphocytes, eosinophils, neutrophils, macrophages, plasma cells, fibroblasts, and collagen. The HRS cell interacts with its microenvironment via cell-cell contact and elaboration of growth factors and cytokines, which results in a surrounding cellular milieu that protects it from host immune attack. The surrounding environmental cells likewise support the HRS cells via cell-cell signaling and cytokine production, which provides signals that promote proliferation and survival of the HRS cell itself. Interestingly, 97% of HRS cells in cHL harbor genetic aberrations in the PD-L1 locus on chromosome 9p24.1, resulting in overexpression of PD-L1, the ligand for the inhibitory PD-1 receptor on immune cells. This is one mechanism whereby the HRS cell may be able to avoid immune destruction in its inflammatory microenvironment and may contribute to the generalized immune suppression in HL patients.

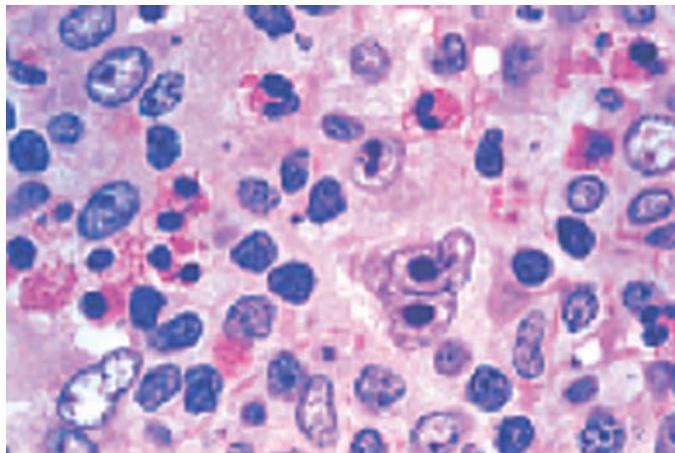


FIGURE 109-1 Hodgkin's disease: A classic Reed-Sternberg (RS) cell is present near the center of the field. RS cells are large cells with a bilobed nucleus and prominent nucleoli surrounded by a pleiomorphic cellular infiltrate. (From DL Kasper: *Harrison's Principles of Internal Medicine*, 16th ed. New York, NY: McGraw-Hill; 2005, Fig. 97-11, p. 654.)

APPROACH TO THE PATIENT

Classical Hodgkin's Lymphoma

Most patients with cHL present with palpable lymphadenopathy that is nontender; in most patients, these lymph nodes are in the neck, supraclavicular area, and axilla. More than half of the patients will have mediastinal adenopathy at diagnosis, and this is sometimes the initial manifestation. Subdiaphragmatic presentation of cHL is unusual and more common in older males. One-third of patients present with fevers, night sweats, and/or weight loss, or “B” symptoms. Occasionally, HL can present as a fever of unknown origin. This is more common in older patients who are found to have mixed-cellularity HL in an abdominal site. Rarely, the fevers persist for days to weeks, followed by afebrile intervals and then recurrence of the fever. This pattern is known as *Pel-Ebstein* fever. HL can occasionally present with unusual manifestations. These include severe and unexplained itching, cutaneous disorders such as erythema nodosum and ichthyosiform atrophy, paraneoplastic cerebellar degeneration and other distant effects on the CNS, nephrotic syndrome, immune hemolytic anemia and thrombocytopenia, hypercalcemia, and pain in lymph nodes on alcohol ingestion.

Evaluation of patients with HL will typically begin with a careful history and physical examination. Patients should be asked about the presence or absence of “B” symptoms. Comorbid diagnoses that may impact therapy should be reviewed, including a history of pulmonary disease and congestive heart failure given the use of chemotherapy drugs that can cause both lung and heart toxicity. A physical examination should pay attention to the peripherally accessible sites of lymph nodes and to the liver and spleen size. Laboratory evaluation should include a complete blood count with differential; erythrocyte sedimentation rate (ESR); chemistry studies reflecting major organ function including serum albumin; and HIV and hepatitis virus testing. A positron emission tomography (PET)/computed tomography (CT) scan is used for staging and is more accurate than a bone marrow biopsy for evaluation of bone marrow involvement as the bone marrow involvement in cHL tends to be patchy and therefore potentially missed on a unilateral bone marrow biopsy. The initial evaluation of a patient with HL or NHL is similar. In both situations, the determination of an accurate anatomic stage is an important part of the evaluation. Staging is done using the Ann Arbor staging system (**Table 109-2**).

The diagnosis of HL is established by review of an adequate biopsy specimen by an expert hematopathologist. HL is a tumor characterized by rare neoplastic cells of B-cell origin (immunoglobulin genes are rearranged but not expressed) in a tumor mass that is largely polyclonal inflammatory infiltrate, probably a reaction to cytokines produced by the tumor cells. The differential diagnosis of a lymph node biopsy suspicious for HL includes inflammatory processes, mononucleosis, NHL, phenytoin-induced adenopathy, and nonlymphomatous malignancies.

Staging for cHL is anatomically based given the propensity of the disease to march from one lymph node group to the next group, often contiguous to the first. Staging is important for selecting therapy of appropriate intensity, but the outcome of optimal therapy for all the stages is excellent. Patients are stratified based on whether they have early-stage disease (stage I or II) or advanced-stage disease (stage III or IV). Patients with early-stage disease have a better prognosis overall but are further classified as favorable or unfavorable based on a variety of factors. These factors vary from study to study but include bulky disease, number of lymph node areas involved, an elevated ESR (>30 if “B” symptoms are present; >50 if “B” symptoms are absent), and age. Prognosis in advanced-stage disease is best predicted by the International Prognostic Score (IPS), which ascribes 1 point for male sex, older age (>45 years), stage IV disease, serum albumin <4 g/dL, hemoglobin <10.5 g/dL, white blood cell count \geq 15,000/ μ L, and a lymphocyte count <600/ μ L and/or <8% of white blood cell count. Five-year progression-free survival ranges from 88% for patients with no risk factors to 62% for patients with four or more factors, but very few patients have multiple risk factors.

TREATMENT

Classical Hodgkin's Lymphoma

The overwhelming majority of patients with HL will be cured with either chemotherapy alone or a combination of chemotherapy and radiation therapy. It has long been appreciated that patients with advanced-stage disease do not benefit from the addition of radiation therapy to chemotherapy and are thus treated with chemotherapy alone. For early-stage disease, however, treatment with combined-modality therapy has been associated with a small decrease in risk of relapse but with an increased risk of late toxicity including secondary malignancies, thyroid disease, and premature cardiovascular disease and stroke resulting in minimal or no improvement in long-term survival. Much of this risk can be attributed to radiation therapy. Thus, investigation into the treatment of early-stage HL at present is aimed at trying to maximize treatment outcome without using radiotherapy. This is an area of controversy in the treatment of HL.

EARLY-STAGE DISEASE

The most common chemotherapy regimen used to treat HL in the United States is ABVD (doxorubicin, bleomycin, vinblastine, and dacarbazine). This regimen is given every other week, with each cycle including two treatments. In patients with low-risk, or favorable, disease, the use of four to six cycles of ABVD alone, without radiation therapy, results in progression-free and overall survival rates of 88–92% and 97–100%, respectively, at 5–7 years. This may be associated with a slightly increased risk of relapse when compared with abbreviated chemotherapy (ABVD for four cycles) followed by involved field radiation therapy (30 Gy), but with no difference in overall survival owing to the excellent salvage strategies used for relapsed HL and to the late toxicities seen following radiation therapy to the chest. German studies have examined a very abbreviated chemotherapy regimen (ABVD for two cycles) and low-dose radiation (20 Gy) for particularly good-risk disease with two or fewer lymph node areas involved and found that this was equally effective to standard combined-modality therapy of ABVD for four cycles and 30 Gy of radiation. However, long-term follow-up is not yet available to assess the impact of the lower

TABLE 109-2 The Ann Arbor Staging System for Hodgkin's Lymphoma

STAGE	DEFINITION
I	Involvement of a single lymph node region or lymphoid structure (e.g., spleen, thymus, Waldeyer's ring)
II	Involvement of two or more lymph node regions on the same side of the diaphragm (the mediastinum is a single site; hilar lymph nodes should be considered “lateralized” and, when involved on both sides, constitute stage II disease)
III	Involvement of lymph node regions or lymphoid structures on both sides of the diaphragm
III ₁	Subdiaphragmatic involvement limited to spleen, splenic hilar nodes, celiac nodes, or portal nodes
III ₂	Subdiaphragmatic involvement includes paraaortic, iliac, or mesenteric nodes plus structures in III ₁
IV	Involvement of extranodal site(s) beyond that designated as “E” More than one extranodal deposit at any location Any involvement of liver or bone marrow
A	No symptoms
B	Unexplained weight loss of >10% of the body weight during the 6 months before staging investigation Unexplained, persistent, or recurrent fever with temperatures >38°C during the previous month Recurrent drenching night sweats during the previous month
E	Localized, solitary involvement of extralymphatic tissue, excluding liver and bone marrow

radiotherapy dose on late toxicities. Finally, the use of an early interim PET/CT scan can aid decisions regarding the duration and extent of therapy. In one study, a negative PET/CT scan after three cycles of ABVD predicted for excellent outcomes with no additional therapy; in another, a negative PET/CT scan after two cycles of ABVD predicted for good outcomes with two additional cycles of ABVD alone, without radiation therapy.

For unfavorable-risk disease, the omission of radiation therapy following chemotherapy is associated with a more significant increased risk of relapse compared to favorable-risk disease, but again with no change in overall survival. For these patients, treatment options would include ABVD for four cycles followed by involved field radiation therapy or ABVD alone for six cycles. Treatment decisions are often based on the extent of the radiation field and the unfavorable risk factor, with patients with nonbulky disease being candidates for chemotherapy alone if radiation would be contraindicated for another reason. Combined modality therapy has typically been used for patients with bulky disease, although patients with bulky disease who have a negative PET/CT scan after chemotherapy may not benefit from additional radiation therapy.

Alternative chemotherapy regimens to ABVD have been developed and include the Stanford V regimen and escalated BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, and prednisone). Neither of these regimens has resulted in improved outcomes in patients with early-stage disease.

ADVANCED-STAGE DISEASE

Patients with advanced-stage disease do not benefit from the addition of radiation therapy after a complete response to chemotherapy alone and should be treated with chemotherapy alone. The most common regimen used in the United States is ABVD for six cycles. Again, Stanford V and escalated BEACOPP have been evaluated in advanced-stage disease and are not associated with an improvement in overall survival but are associated with increased toxicity. The small fraction of patients who do not achieve complete remission with chemotherapy alone (partial responders with persistent PET scan positivity account for <10% of patients) may benefit from the addition of involved field radiotherapy.

Newer drugs have been developed for the treatment of relapsed HL (see “Relapsed Disease,” below). These include the antibody-drug conjugate brentuximab vedotin, which is an antibody against CD30 conjugated to the microtubule inhibitor monomethyl auristatin E (MMAE). This drug has been combined with doxorubicin, bleomycin, and dacarbazine in early-phase studies for advanced-stage HL with favorable efficacy compared to historical controls. Eschelon-1, a randomized trial of doxorubicin, vinblastine, and dacarbazine (AVD) plus brentuximab compared to ABVD, was a positive study in that it demonstrated an improvement in progression-free survival for AVD plus brentuximab, especially among younger patients, patients from North America, and patients with higher risk disease. Drugs that target the PD-1/PD-L1 axis have been developed in an attempt to boost the host immune recognition of tumors. This was particularly attractive in HL given the overexpression of PD-L1 on the HRS cell surface. In the setting of relapsed disease, these drugs, which include pembrolizumab and nivolumab, have very high response rates and are associated with durable responses. These are now being tested in conjunction with chemotherapy both as salvage therapy for relapsed disease and in previously untreated patients, including in a multicenter randomized trial against AVD plus brentuximab as initial therapy for advanced-stage disease.

RELAPSED DISEASE

Patients who relapse after primary therapy of HL can frequently still be cured. Patients who relapse after an effective chemotherapy regimen are usually not curable with subsequent chemotherapy administered at standard doses. Alternative salvage chemotherapy administered at standard doses, then, is given in order to document sensitivity to chemotherapy and to achieve maximum reduction of tumor mass. For patients who respond completely or nearly so, autologous stem cell transplantation can cure over

half of patients. Standard salvage chemotherapy regimens include ICE (ifosfamide, carboplatin, and etoposide) and GND (gemcitabine, vinorelbine, and doxorubicin). Newer combinations, including brentuximab with either chemotherapy or immune checkpoint inhibitors such as nivolumab, have also been tested with promising early results. For patients with early-stage disease who do not respond sufficiently to salvage chemotherapy, radiation therapy can be very effective to achieve a remission; whether to consolidate such a remission with an autologous stem cell transplant is debated. For patients with advanced-stage disease in whom salvage chemotherapy fails, the antibody-drug conjugate brentuximab vedotin, a CD30-directed antibody linked to the microtubule toxin MMAE, is active and can be tried as a bridge to allogeneic transplant. It is also used as a maintenance therapy following successful autologous stem cell transplantation based on results of a randomized trial versus observation. The anti-PD-1 immune checkpoint inhibitors, nivolumab and pembrolizumab, have efficacy in relapsed HL, and many responses are durable. Increasingly, there is an appreciation that use of checkpoint inhibitors restores the HRS cell’s sensitivity to chemotherapy by unknown mechanisms; autologous stem cell transplantation may be a potentially curative option for patients who had previously been felt to have chemotherapy-resistant disease. Finally, anti-CD30 chimeric antigen receptor (CAR) T-cell therapy has been tested in multiply relapsed cHL with promising early results; these products are now being tested in multicenter phase 2 clinical trials.

Two other options may be useful in the setting of disease relapse after ABVD chemotherapy. Alkylating agent-based combinations such as ChlVPP (chlorambucil, vincristine, prednisone, and procarbazine) may be active in patients with disease resistant to ABVD. In addition, relapse following bone marrow transplant can be responsive to weekly low-dose single-agent vinblastine.

SURVIVORSHIP

Because of the very high cure rate in patients with HL, long-term complications have become a major focus for clinical research. In fact, in some series of patients with early-stage disease, more patients died from late complications of therapy than from HL itself. This is particularly true in patients with localized disease. The most serious late side effects include second malignancies and cardiac injury. Patients are at risk for the development of acute leukemia in the first 10 years after treatment with combination chemotherapy regimens that contain alkylating agents plus radiation therapy. The risk for development of acute leukemia is greater after MOPP-like (mechlorethamine, vincristine, procarbazine, and prednisone) and BEACOPP-like regimens than with ABVD. The risk of development of acute leukemia after treatment for HL is also related to the number of exposures to potentially leukemogenic agents (i.e., multiple treatments after relapse) and the age of the patient being treated, with those aged >60 years at particularly high risk. The development of carcinomas as a complication of treatment for HL is a major problem. These tumors usually occur ≥10 years after treatment and are associated with use of radiotherapy. For this reason, young women treated with thoracic radiotherapy for HL should institute screening mammograms 5–10 years after treatment, and all patients who receive thoracic radiotherapy for HL should be discouraged from smoking. Mediastinal radiation also accelerates coronary artery disease, and patients should be encouraged to minimize risk factors for coronary artery disease such as smoking and elevated cholesterol levels. Cervical radiation therapy increases the risk of carotid atherosclerosis and stroke and thyroid disease, including cancer.

A number of other late side effects from the treatment of HL are well known. Patients who receive thoracic radiotherapy are at very high risk for the eventual development of hypothyroidism and should be observed for this complication; intermittent measurement of thyrotropin should be made to identify the condition before it becomes symptomatic. Lhermitte’s syndrome occurs in ~15% of patients who receive thoracic radiotherapy. This syndrome is manifested by an “electric shock” sensation into the lower extremities on flexion of the neck. Because of the young age at which HL is often diagnosed, infertility is a concern for patients undergoing treatment for HL. Chemotherapy

regimens containing alkylating agents induce permanent infertility in nearly all men. The risk of permanent infertility in women treated with alkylating agent-containing chemotherapy is age-related, with younger women more likely to recover fertility. Infertility is very rare after treatment with ABVD.

NODULAR LYMPHOCYTE-PREDOMINANT HODGKIN'S LYMPHOMA

NLPHL is now recognized as an entity distinct from cHL. Previous classification systems recognized that biopsies from a small subset of patients diagnosed as having HL contained a predominance of small lymphocytes and rare Reed-Sternberg-like cells; tumors had a nodular growth pattern and a clinical course that varied from that of patients with cHL. This is an unusual clinical entity and represents <5% of cases of HL and defines NLPHL.

NLPHL has a number of characteristics that suggest its relationship to NHL, rather than cHL, however. The HRS-like cell, or L&H (lymphocyte and histiocyte) or “popcorn” cell, is a clonal proliferation of B-cells that are positive for B-cell markers CD45, CD79a, CD20, CD19, and BCL2. They do not express two markers normally found on HRS cells, CD30 and CD15. This lymphoma tends to have a chronic, relapsing course and sometimes transforms to diffuse large B-cell lymphoma, including a specific subtype of diffuse large B-cell lymphoma known as T-cell/histiocyte-rich B-cell lymphoma, which shares an immunophenotype with the L&H cell. This natural history most closely resembles that of the indolent B-cell NHLs outlined in Chaps. 108 and 110.

Patients with NLPHL are more commonly male (75%). Like cHL, the age distribution of patients with this disease has two peaks, but unlike cHL, these peaks include children and adults ages 30–40 years, respectively. The majority of patients diagnosed have stage I or II disease (75%), with a minority having advanced-stage disease at diagnosis. “B” symptoms are uncommon.

Patients with early-stage disease at diagnosis should be treated with definitive radiotherapy. This is associated with a 15-year nonrelapse survival rate of 82%. The treatment of patients with advanced-stage NLPHL is controversial. Some clinicians favor no treatment of asymptomatic disease and merely close follow-up, akin to the indolent B-cell NHLs. For patients who need therapy due to symptoms or signs of organ function impairment, both cHL regimens and B-cell NHL regimens have been used, including ABVD and R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone). A single-institution experience with R-CHOP resulted in a 100% response rate in a small group of patients without a single relapse with 42 months of follow-up. Although this is short follow-up for an indolent disease, some believe R-CHOP may be curative in this disease and advocate treating patients with advanced-stage disease at diagnosis, regardless of symptoms or organ function.

FURTHER READING

- CHEN R et al: Pembrolizumab in relapsed or refractory Hodgkin lymphoma: 2-year follow-up of KEYNOTE-087. *Blood* 134:1144, 2019.
- GILLESSEN S et al: Intensified treatment of patients with early stage, unfavourable Hodgkin lymphoma: Long-term follow-up of a randomised, international phase 3 trial of the German Hodgkin Study Group (GHSG HD14). *Lancet Haematol* 8:e278, 2021.
- MOSKOWITZ CH et al: Five-year PFS from the AETHERA trial of brentuximab vedotin for Hodgkin lymphoma at high risk of progression or relapse. *Blood* 132:2639, 2018.
- RASHIDI A et al: Allogeneic hematopoietic stem cell transplantation in Hodgkin lymphoma: A systemic review and meta-analysis. *Bone Marrow Transplant* 51:521, 2016.
- STRAUS DJ et al: CALGB 50604: Risk-adapted treatment of non-bulky early-stage Hodgkin lymphoma based on interim PET. *Blood* 132:1013, 2018.
- STRAUS DJ et al: Brentuximab vedotin with chemotherapy for stage III or IV classical Hodgkin lymphoma (ECHELON-1): 5-year update of an international, open-label, randomised, phase 3 trial. *Lancet Haematol* 8:e410, 2021.

110

Less Common Lymphoid and Myeloid Malignancies

Ayalew Tefferi, Dan L. Longo

The most common lymphoid malignancies are discussed in Chaps. 106, 107, 108, 109, and 111, myeloid leukemias in Chaps. 104 and 105, myelodysplastic syndromes (MDS) in Chap. 102, and myeloproliferative syndromes in Chap. 103. This chapter will focus on the more unusual forms of hematologic malignancy. The diseases discussed here are listed in Table 110-1. Each of these entities accounts for <1% of hematologic neoplasms.

RARE LYMPHOID MALIGNANCIES

All the lymphoid tumors discussed here are mature B-cell or T-cell natural killer (NK) cell neoplasms.

MATURE B-CELL NEOPLASMS

B-Cell Prolymphocytic Leukemia (B-PLL) This is a malignancy of medium-sized (about twice the size of a normal small

TABLE 110-1 Unusual Lymphoid and Myeloid Malignancies

Lymphoid
Mature B-cell neoplasms
B-cell prolymphocytic leukemia
Splenic marginal zone lymphoma
Hairy cell leukemia
Nodal marginal zone B-cell lymphoma
Mediastinal large B-cell lymphoma
Intravascular large B-cell lymphoma
Primary effusion lymphoma
Lymphomatoid granulomatosis
Mature T-cell and natural killer (NK) cell neoplasms
T-cell prolymphocytic leukemia
T-cell large granular lymphocytic leukemia
Aggressive NK cell leukemia
Extranodal NK/T-cell lymphoma, nasal type
Enteropathy-type T-cell lymphoma
Hepatosplenitic T-cell lymphoma
Subcutaneous panniculitis-like T-cell lymphoma
Blastic NK cell lymphoma
Primary cutaneous CD30+ T-cell lymphoma
Angioimmunoblastic T-cell lymphoma
Myeloid
Chronic neutrophilic leukemia
Chronic eosinophilic leukemia/hypereosinophilic syndrome
Histiocytic and Dendritic Cell Neoplasms
Histiocytic sarcoma
Langerhans cell histiocytosis
Langerhans cell sarcoma
Interdigitating dendritic cell sarcoma
Follicular dendritic cell sarcoma
Mast Cells
Mastocytosis
Cutaneous mastocytosis
Systemic mastocytosis
Mast cell sarcoma
Extracutaneous mastocytoma

lymphocyte), round lymphocytes with a prominent nucleolus and light blue cytoplasm on Wright's stain. It predominantly affects the blood, bone marrow (BM), and spleen and usually does not cause adenopathy. The median age of affected patients is 70 years, and men are more often affected than women (male-to-female ratio is 1.6). This entity is distinct from chronic lymphoid leukemia (CLL) and does not develop as a consequence of that disease.

Clinical presentation is generally from symptoms of splenomegaly or incidental detection of an elevated white blood cell (WBC) count. The clinical course can be rapid. The cells express surface IgM (with or without IgD) and typical B-cell markers (CD19, CD20, CD22). CD23 is absent, and about one-third of cases express CD5. The CD5 expression along with the presence of the t(11;14) translocation in 20% of cases leads to confusion in distinguishing B-PLL from the leukemic form of mantle cell lymphoma. No reliable criteria for the distinction have emerged, and gene expression studies suggest a close relationship between mantle cell lymphoma and B-PLL and significant differences with CLL. About half of patients have mutation or loss of p53, and deletions have been noted in 11q23 and 13q14. Nucleoside analogues like fludarabine and cladribine and combination chemotherapy (cyclophosphamide, doxorubicin, vincristine, and prednisone [CHOP]) have produced responses. CHOP plus rituximab may be more effective than CHOP alone, but the disease is sufficiently rare that large series have not been reported. Splenectomy can produce palliation of symptoms but appears to have little or no impact on the course of the disease. BM transplantation may be curative. Imatinib may also have activity.

Splenic Marginal Zone Lymphoma (SMZL) This tumor of mainly small lymphocytes originates in the marginal zone of the spleen white pulp, grows to efface the germinal centers and mantle, and invades the red pulp. Splenic hilar nodes, BM, and peripheral blood (PB) may be involved. The circulating tumor cells have short surface villi and are called villous lymphocytes. **Table 110-2** shows differences in tumor cells of a number of neoplasms of small lymphocytes that aid in the differential diagnosis. SMZL cells express surface immunoglobulin and CD20 but are negative for CD5, CD10, CD43, and CD103. Lack of CD5 distinguishes SMZL from CLL, and lack of CD103 separates SMZL from hairy cell leukemia.

The median age of patients with SMZL is mid-fifties, and men and women are equally represented. Patients present with incidental or symptomatic splenomegaly or incidental detection of lymphocytosis in the PB with villous lymphocytes. Autoimmune anemia or thrombocytopenia may be present. The immunoglobulin produced by these cells contains somatic mutations that reflect transit through a germinal center, and ongoing mutations suggest that the mutation machinery has remained active. About 40% of patients have either deletions or translocations involving 7q21, the site of the *FLNC* gene (filamin C_y, involved in cross-linking actin filaments in the cytoplasm). *NOTCH2* mutations are seen in 25% of patients. Chromosome 8p deletions may

TABLE 110-3 Differential Diagnosis of “Dry Tap”—Inability to Aspirate Bone Marrow

Dry taps occur in about 4% of attempts and are associated with:

Metastatic carcinoma infiltration	17%
Chronic myeloid leukemia	15%
Myelofibrosis	14%
Hairy cell leukemia	10%
Acute leukemia	10%
Lymphomas, Hodgkin’s disease	9%
Normal marrow	Rare

also be noted. The genetic lesions typically found in extranodal marginal zone lymphomas (e.g., trisomy 3 and t[11;18]) are uncommon in SMZL.

The clinical course of disease is generally indolent with median survivals exceeding 10 years. Patients with elevated lactate dehydrogenase (LDH) levels, anemia, and hypoalbuminemia generally have a poorer prognosis. Long remissions can be seen after splenectomy. Rituximab, ibrutinib, and PI3 kinase inhibitors are also active. A small fraction of patients undergo histologic progression to diffuse large B-cell lymphoma with a concomitant change to a more aggressive natural history. Experience with combination chemotherapy in SMZL is limited.

Hairy Cell Leukemia Hairy cell leukemia is a tumor of small lymphocytes with oval nuclei, abundant cytoplasm, and distinctive membrane projections (hairy cells). Patients have splenomegaly and diffuse BM involvement. While some circulating cells are noted, the clinical picture is dominated by symptoms from the enlarged spleen and pancytopenia. The mechanism of the pancytopenia is not completely clear and may be mediated by both inhibitory cytokines and marrow replacement. The marrow has an increased level of reticulin fibers; indeed, hairy cell leukemia is a common cause of inability to aspirate BM or so-called “dry tap” (**Table 110-3**). Monocytopenia is profound and may explain a predisposition to atypical mycobacterial infection that is observed clinically. The tumor cells have strong expression of CD22, CD25, and CD103; soluble CD25 level in serum is an excellent tumor marker for disease activity. The cells also express tartrate-resistant acid phosphatase. The immunoglobulin genes are rearranged and mutated, indicating the influence of a germinal center. No specific cytogenetic abnormality has been found, but most cases contain the activating *BRAF* mutation V600E.

The median age of affected patients is mid-fifties, and the male-to-female ratio is 5:1. Treatment options are numerous. Splenectomy is often associated with prolonged remission. Nucleosides including cladribine and deoxycoformycin are highly active but are also associated with further immunosuppression and can increase the risk of certain opportunistic infections. However, after brief courses of these agents, patients usually obtain very durable remissions during which immune function spontaneously recovers. Interferon α is also an effective therapy but is not as effective as nucleosides. Chemotherapy-refractory patients have responded to vemurafenib, a BRAF inhibitor. Vemurafenib does not appear to be curative, but responses can be maintained with chronic treatment. More durable remissions occur when rituximab is added to vemurafenib.

Nodal Marginal Zone B Cell Lymphoma This rare node-based disease bears an uncertain relationship with extranodal marginal zone lymphomas, which are often mucosa-associated and are called mucosa-associated lymphoid tissue (MALT) lymphomas, and SMZLs. Patients may have localized or generalized adenopathy. The neoplastic cell is a marginal zone B cell with monocyteoid features and has been called monocyteoid B cell lymphoma in the past. Up to one-third of the patients may have extranodal involvement, and involvement of the lymph nodes can be secondary to the spread of a mucosal primary lesion. In authentic nodal primaries, the cytogenetic abnormalities associated with MALT lymphomas (trisomy 3 and t[11;18]) are very rare. The clinical course is indolent. Patients often respond

TABLE 110-2 Immunophenotype of Tumors of Small Lymphocytes

	CD5	CD20	CD43	CD10	CD103	sIG	CYCLIN D1
Follicular lymphoma	neg	pos	pos	pos	neg	pos	neg
Chronic lymphoid leukemia	pos	pos	pos	neg	neg	pos	neg
B-cell prolymphocytic leukemia	pos	pos	pos	neg	neg	pos	pos
Mantle cell lymphoma	pos	pos	pos	neg	neg	pos	pos
Splenic marginal zone lymphoma	neg	pos	neg	neg	neg	pos	neg
Hairy cell leukemia	neg	pos	?	neg	pos	pos	neg

Abbreviations: neg, negative; pos, positive.

to combination chemotherapy, although remissions have not been durable. Few patients have received CHOP plus rituximab, which is likely to be an effective approach to management.

Mediastinal (Thymic) Large B-Cell Lymphoma This entity was originally considered a subset of diffuse large B-cell lymphoma; however, additional study has identified it as a distinct entity with its own characteristic clinical, genetic, and immunophenotypic features. This is a disease that can be bulky in size but usually remains confined to the mediastinum. It can be locally aggressive, including progressing to produce a superior vena cava obstruction syndrome or pericardial effusion. About one-third of patients develop pleural effusions, and in 5–10% of cases, disease can disseminate widely to kidney, adrenal, liver, skin, and even brain. The disease affects women more often than men (male-to-female ratio is 1:2–3), and the median age is 35–40 years.

The tumor is composed of sheets of large cells with abundant cytoplasm accompanied by variable, but often abundant, fibrosis. It is distinguished from nodular sclerosing Hodgkin's disease by the paucity of normal lymphoid cells and the absence of lacunar variants of Reed-Sternberg cells. However, more than one-third of the genes that are expressed to a greater extent in primary mediastinal large B-cell lymphoma than in usual diffuse large B-cell lymphoma are also overexpressed in Hodgkin's disease, suggesting a possible pathogenetic relationship between the two entities that affect the same anatomic site. Tumor cells may overexpress *MAL*. The genome of tumor cells is characterized by frequent chromosomal gains and losses. The tumor cells in mediastinal large B-cell lymphoma express CD20, but surface immunoglobulin and human leukocyte antigen (HLA) class I and class II molecules may be absent or incompletely expressed. Expression of lower levels of class II HLA identifies a subset with poorer prognosis. The cells are CD5 and CD10 negative but may show light staining with anti-CD30. The cells are CD45 positive, unlike cells of classical Hodgkin's disease.

Methotrexate, leucovorin, doxorubicin, cyclophosphamide, vinristine, prednisone, and bleomycin (MACOP-B) and rituximab plus CHOP are effective treatments, achieving 5-year survival of 75–87%. Dose-adjusted therapy with prednisone, etoposide, vincristine, cyclophosphamide, and doxorubicin (EPOCH) plus rituximab has produced 5-year survival of 97%. A role for mediastinal radiation therapy has not been definitively demonstrated, but it is frequently used, especially in patients whose mediastinal area remains positron emission tomography-avid after 4–6 cycles of chemotherapy.

Intravascular Large B-Cell Lymphoma This is an extremely rare form of diffuse large B-cell lymphoma characterized by the presence of lymphoma in the lumen of small vessels, particularly capillaries. It is also known as malignant angioendotheliomatosis or angiotropic large-cell lymphoma. It is sufficiently rare that no consistent picture has emerged to define a clinical syndrome or its epidemiologic and genetic features. It is thought to remain inside vessels because of a defect in adhesion molecules and homing mechanisms, an idea supported by scant data suggesting absence of expression of β -1 integrin and ICAM-1. Patients commonly present with symptoms of small-vessel occlusion, skin lesions, or neurologic symptoms. The tumor cell clusters can promote thrombus formation. A subset of patients have tumors with *MYD88* or *CD79B* mutations. In general, the clinical course is aggressive and the disease is poorly responsive to therapy. Often a diagnosis is not made until very late in the course of the disease or at autopsy.

Primary Effusion Lymphoma This entity is another variant of diffuse large B-cell lymphoma that presents with pleural effusions, usually without apparent tumor mass lesions. It is most common in the setting of immune deficiency disease, especially AIDS, and is caused by human herpes virus 8 (HHV-8)/Kaposi's sarcoma herpes virus (KSHV). It is also known as *body cavity-based lymphoma*. Some patients have been previously diagnosed with Kaposi's sarcoma. It can also occur in the absence of immunodeficiency in elderly men of Mediterranean heritage, similar to Kaposi's sarcoma but even less common.

The malignant effusions contain cells positive for HHV-8/KSHV, and many are also co-infected with Epstein-Barr virus. The cells

are large with large nuclei and prominent nucleoli that can be confused with Reed-Sternberg cells. The cells express CD20 and CD79a (immunoglobulin-signaling molecule), although they often do not express immunoglobulin. Some cases aberrantly express T-cell markers such as CD3 or rearranged T-cell receptor genes. No characteristic genetic lesions have been reported, but gains in chromosome 12 and X material have been seen, similar to other HIV-associated lymphomas. The clinical course is generally characterized by rapid progression and death within 6 months. CHOP plus lenalidomide or bortezomib may produce responses. Highly active antiretroviral therapy for HIV should be maintained during treatment.

Lymphomatoid Granulomatosis This is an angiocentric, angiolytic lymphoproliferative disease comprised by neoplastic Epstein-Barr virus-infected monoclonal B cells accompanied and outnumbered by a polyclonal reactive T-cell infiltrate. The disease is graded based on histologic features such as cell number and atypia in the B cells. It is most often confused with extranodal NK/T-cell lymphoma, nasal type, which can also be angiolytic and is Epstein-Barr virus-related. The disease usually presents in adults (males > females) as a pulmonary infiltrate. Involvement is often entirely extranodal and can include kidney (32%), liver (29%), skin (25%), and brain (25%). The disease often but not always occurs in the setting of immune deficiency.

The disease can be remitting and relapsing in nature or can be rapidly progressive. The course is usually predicted by the histologic grade. The disease is highly responsive to combination chemotherapy and is curable in most cases. Some investigators have claimed that low-grade disease (grade I and II) can be treated with interferon α .

■ MATURE T-CELL AND NK CELL NEOPLASMS

T-Cell Prolymphocytic Leukemia This is an aggressive leukemia of medium-sized prolymphocytes involving the blood, marrow, nodes, liver, spleen, and skin. It accounts for 1–2% of all small lymphocytic leukemias. Most patients present with elevated WBC count (often >100,000/ μ L), hepatosplenomegaly, and adenopathy. Skin involvement occurs in 20%. The diagnosis is made from PB smear, which shows cells about 25% larger than those in small lymphocytes, with cytoplasmic blebs and nuclei that may be indented. The cells express T-cell markers like CD2, CD3, and CD7; two-thirds of patients have cells that are CD4+ and CD8-, and 25% have cells that are CD4+ and CD8+. T-cell receptor β chains are clonally rearranged. In 80% of patients, inversion of chromosome 14 occurs between q11 and q32. Ten percent have t(14;14) translocations that bring the T-cell receptor alpha/beta gene locus into juxtaposition with oncogenes *TCL1* and *TCL1b* at 14q32.1. Chromosome 8 abnormalities are also common. Deletions in the *ATM* gene are also noted. Activating *JAK3* mutations have also been reported.

The course of the disease is generally rapid, with median survival of about 12 months. Responses have been seen with the anti-CD52 antibody alemtuzumab, nucleoside analogues, and CHOP chemotherapy. Histone deacetylase inhibitors like vorinostat and romidepsin may also have activity. Small numbers of patients with T-cell prolymphocytic leukemia have also been treated with high-dose therapy, and allogeneic BM transplantation after remission has been achieved with alemtuzumab or conventional-dose therapy.

T-Cell Large Granular Lymphocytic Leukemia T-cell large granular lymphocytic (LGL) leukemia is characterized by increases in the number of LGLs in the PB (2000–20,000/ μ L) often accompanied by severe neutropenia, with or without concomitant anemia. Patients may have splenomegaly and frequently have evidence of systemic autoimmune disease, including rheumatoid arthritis, hypergammaglobulinemia, autoantibodies, and circulating immune complexes. BM involvement is mainly interstitial in pattern, with <50% lymphocytes on differential count. Usually the cells express CD3, T-cell receptors, and CD8; NK-like variants may be CD3-. The leukemic cells often express Fas and Fas ligand.

The course of the disease is generally indolent and dominated by the neutropenia. Paradoxically, immunosuppressive therapy with cyclosporine, methotrexate, or cyclophosphamide plus glucocorticoids can produce an increase in granulocyte counts. Nucleosides have been used anecdotally. Occasionally the disease can accelerate to a more aggressive clinical course.

Aggressive NK Cell Leukemia NK neoplasms are very rare, and they may follow a range of clinical courses from very indolent to highly aggressive. They are more common in Asians than whites, and the cells frequently harbor a clonal Epstein-Barr virus episome. The PB white count is usually not greatly elevated, but abnormal large lymphoid cells with granular cytoplasm are noted. The aggressive form is characterized by symptoms of fever and laboratory abnormalities of pancytopenia. Hepatosplenomegaly is common; node involvement is less common. Patients may have hemophagocytosis, coagulopathy, or multiorgan failure. Serum levels of Fas ligand are elevated.

The cells express CD2 and CD56 and do not have rearranged T-cell receptor genes. Deletions involving chromosome 6 are common. The disease can be rapidly progressive. Some forms of NK neoplasms are more indolent. They tend to be discovered incidentally with LGL lymphocytosis and do not manifest the fever and hepatosplenomegaly characteristic of the aggressive leukemia. The cells are also CD2 and CD56 positive, but they do not contain clonal forms of Epstein-Barr virus and are not accompanied by pancytopenia or autoimmune disease.

Extranodal NK/T-Cell Lymphoma, Nasal Type Like lymphomatoid granulomatosis, extranodal NK/T-cell lymphoma tends to be an angiocentric and angiolytic lesion, but the malignant cells are not B cells. In most cases, they are CD56+ Epstein-Barr virus-infected cells; occasionally, they are CD56-Epstein-Barr virus-infected cytotoxic T cells. They are most commonly found in the nasal cavity. Historically, this illness was called lethal midline granuloma, polymorphic reticulosis, and angiocentric immunoproliferative lesion. This form of lymphoma is prevalent in Asia, Mexico, and Central and South America; it affects males more commonly than females. When it spreads beyond the nasal cavity, it may affect soft tissue, the gastrointestinal tract, or the testis. In some cases, hemophagocytic syndrome (HPS) may influence the clinical picture. Patients may have B symptoms. Many of the systemic manifestations of disease are related to the production of cytokines by the tumor cells and the cells responding to their signals. Deletions and inversions of chromosome 6 are common.

Many patients with extranodal NK/T-cell lymphoma, nasal type, have excellent antitumor responses with combination chemotherapy regimens, particularly those with localized disease. Radiation therapy is often used after completion of chemotherapy. Four risk factors have been defined, including B symptoms, advanced stage, elevated LDH, and regional lymph node involvement. Patient survival is linked to the number of risk factors: 5-year survival is 81% for zero risk factors, 64% for one risk factor, 32% for two risk factors, and 7% for three or four risk factors. Combination regimens without anthracyclines have been touted as superior to CHOP, but data are sparse. High-dose therapy with stem cell transplantation has been used, but its role is unclear.

Enteropathy-Type T-Cell Lymphoma Enteropathy-type T-cell lymphoma is a rare complication of longstanding celiac disease. It most commonly occurs in the jejunum or the ileum. In adults, the lymphoma may be diagnosed at the same time as celiac disease, but the suspicion is that the celiac disease was a longstanding precursor to the development of lymphoma. The tumor usually presents as multiple ulcerating mucosal masses, but may also produce a dominant exophytic mass or multiple ulcerations. The tumor expresses CD3 and CD7 nearly always and may or may not express CD8. The normal-appearing lymphocytes in the adjacent mucosa often have a similar phenotype to the tumor. Most patients have the HLA genotype associated with celiac disease, HLA DQA1*0501 or DQB1*0201.

The prognosis of this form of lymphoma is typically poor (median survival is 7 months), but some patients have a good response to CHOP chemotherapy. Patients who respond can develop bowel perforation

from responding tumor. If the tumor responds to treatment, recurrence may develop elsewhere in the celiac disease-affected small bowel.

Hepatosplenic T-Cell Lymphoma Hepatosplenic T-cell lymphoma is a malignancy derived from T cells expressing the gamma/delta T-cell antigen receptor that affects mainly the liver and fills the sinusoids with medium-size lymphoid cells. When the spleen is involved, dominantly the red pulp is infiltrated. It is a disease of young people, especially young people with an underlying immunodeficiency or with an autoimmune disease that demands immunosuppressive therapy. The use of thiopurine and infliximab is particularly common in the history of patients with this disease. The cells are CD3+ and usually CD4- and CD8-. The cells may contain isochromosome 7q, often together with trisomy 8. The lymphoma has an aggressive natural history. Combination chemotherapy may induce remissions, but most patients relapse. Median survival is about 2 years. The tumor does not appear to respond to reversal of immunosuppressive therapy.

Subcutaneous Panniculitis-Like T-Cell Lymphoma Subcutaneous panniculitis-like T-cell lymphoma involves multiple subcutaneous collections of neoplastic T cells that are usually cytotoxic cells in phenotype (i.e., contain perforin and granzyme B and express CD3 and CD8). The rearranged T-cell receptor is usually alpha/beta-derived, but occasionally, the gamma/delta receptors are involved, particularly in the setting of immunosuppression. The cells are negative for Epstein-Barr virus. Patients may have an HPS in addition to the skin infiltration; fever and hepatosplenomegaly may also be present. Nodes are generally not involved. Patients frequently respond to combination chemotherapy, including CHOP. When the disease is progressive, the HPS can be a component of a fulminant downhill course. Effective therapy can reverse the HPS.

Blastic NK Cell Lymphoma The neoplastic cells express NK cell markers, especially CD56, and are CD3 negative. They are large blastic-looking cells and may produce a leukemia picture, but the dominant site of involvement is the skin. Morphologically, the cells are similar to the neoplastic cells in acute lymphoid and myeloid leukemia. No characteristic chromosomal abnormalities have been described. The clinical course is rapid, and the disease is largely unresponsive to typical lymphoma treatments.

Primary Cutaneous CD30+ T-Cell Lymphoma This tumor involves the skin and is composed of cells that appear similar to the cells of anaplastic T-cell lymphoma. Among cutaneous T-cell tumors, ~25% are CD30+ anaplastic lymphomas. If dissemination to lymph nodes occurs, it is difficult to distinguish between the cutaneous and systemic forms of the disease. The tumor cells are often CD4+, and the cells contain granules that are positive for granzyme B and perforin in 70% of cases. The typical t(2;5) of anaplastic T-cell lymphoma is absent; indeed, its presence should prompt a closer look for systemic involvement and a switch to a diagnosis of anaplastic T-cell lymphoma. This form of lymphoma has sporadically been noted as a rare complication of silicone or saline breast implants. The natural history of breast implant-associated lymphoma is generally indolent. Cutaneous CD30+ T-cell lymphoma often responds to therapy. The anti-CD30 immunotoxin conjugate brentuximab vedotin is active. Radiation therapy can be effective, and surgery can also produce long-term disease control. Five-year survival exceeds 90%.

Angioimmunoblastic T-Cell Lymphoma Angioimmunoblastic T-cell lymphoma is a systemic disease that accounts for ~15% of all T-cell lymphomas. Patients frequently have fever, advanced stage, diffuse adenopathy, hepatosplenomegaly, skin rash, polyclonal hypergammaglobulinemia, and a wide range of autoantibodies including cold agglutinins, rheumatoid factor, and circulating immune complexes. Patients may have edema, arthritis, pleural effusions, and ascites. The nodes contain a polymorphous infiltrate of neoplastic T cells and nonneoplastic inflammatory cells together with proliferation of high endothelial venules and follicular dendritic cells (FDCs). The most common chromosomal abnormalities are trisomy 3, trisomy 5, and an extra X chromosome. Aggressive combination chemotherapy can

induce regressions. The underlying immune defects make conventional lymphoma treatments more likely to produce infectious complications.

RARE MYELOID MALIGNANCIES

The World Health Organization (WHO) system uses PB counts and smear analysis, BM morphology, and cytogenetic and molecular genetic tests in order to classify myeloid malignancies into several major categories (**Table 110-4**). Among them, acute myeloid leukemia (AML) is discussed in **Chap. 104**, myelodysplastic syndromes (MDS) in **Chap. 102**, chronic myeloid leukemia (CML) in **Chap. 105**, and JAK2 mutation-enriched myeloproliferative neoplasms (MPNs) in **Chap. 103**. In this chapter, we focus on the rest (listed in Table 110-4) including chronic neutrophilic leukemia (CNL); atypical CML, *BCR-ABL1* negative (aCML); chronic myelomonocytic leukemia (CMML); juvenile myelomonocytic leukemia (JMML); chronic eosinophilic leukemia, not otherwise specified (CEL-NOS); mastocytosis; MPN, unclassifiable (MPN-U); MDS/MPN, unclassifiable (MDS/MPN-U); MDS/MPN with ring sideroblasts and thrombocytosis (MDS/MPN-RS-T); and myeloid/lymphoid neoplasms with eosinophilia and rearrangements of *PDGFRA*, *PDGFRB*, or *FGFR1* or with *PCM1-JAK2*. This chapter also includes histiocytic and dendritic cell neoplasms, transient myeloproliferative disorders, and a broader discussion on primary eosinophilic disorders including hypereosinophilic syndrome (HES).

CHRONIC NEUTROPHILIC LEUKEMIA

CNL is a clonal proliferation of mature neutrophils with few or no circulating immature granulocytes. In 2013, CNL was described to be associated with activating mutations of the gene (*CSF3R*) encoding

for the receptor for granulocyte colony-stimulating factor (G-CSF), also known as colony-stimulating factor 3 (CSF3). Patients with CNL might be asymptomatic at presentation but also display constitutional symptoms, splenomegaly, anemia, and thrombocytopenia. Median survival is approximately 2 years and causes of death include leukemic transformation, progressive disease associated with severe cytopenias and marked treatment-refractory leukocytosis. The true incidence of CNL is not known due to diagnostic uncertainty with >200 currently reported cases. Median age at diagnosis is approximately 67 years with a slight male preponderance in gender distribution.

Pathogenesis CSF3 is the main growth factor for granulocyte proliferation and differentiation. Accordingly, recombinant CSF3 is used for the treatment of severe neutropenia, including severe congenital neutropenia (SCN). Some patients with SCN acquire *CSF3R* mutations, and the frequency of such mutations is significantly higher (~80%) in patients who experience leukemic transformation. SCN-associated *CSF3R* mutations occur in the region of the gene coding for the cytoplasmic domain of *CSF3R* and result in truncation of the C-terminal-negative regulatory domain. In 2013, Maxson et al described a different class of *CSF3R* mutations in ~90% of patients with CNL; these were mostly membrane proximal, the most frequent being a C-to-T substitution at nucleotide 1853 (T618I). In a subsequent confirmatory study, *CSF3R* mutations were found to be specific to WHO-defined CNL. About 40% of the T618I-mutated cases also harbored *SETBP1* mutations. *CSF3R* T618I has been shown to induce lethal myeloproliferative disorder in a mouse model and in vitro sensitivity to JAK inhibition.

Diagnosis Diagnosis of CNL requires exclusion of the more common causes of neutrophilia including infections and inflammatory processes. In addition, one should be mindful of the association between some forms of metastatic cancer or plasma cell neoplasms with secondary neutrophilia. Neoplastic neutrophilia also occurs in other *BCR-ABL1*-negative myeloid malignancies including aCML and CMML. Accordingly, the WHO diagnostic criteria for CNL are designed to exclude the possibilities of both secondary/reactive neutrophilia and leukocytosis associated with myeloid malignancies other than CNL (**Table 110-5**): leukocytosis ($\geq 25 \times 10^9/L$), $\geq 80\%$ segmented/band neutrophils, $<10\%$ immature myeloid cells, $<1\%$ circulating blasts, and absence of dysgranulopoiesis or monocytosis (monocyte count $<1 \times 10^9/L$). BM in CNL is hypercellular and displays increased number and percentage of neutrophils with a very high myeloid-to-erythroid ratio and minimal left shift, myeloid dysplasia, or reticulin fibrosis.

The recent discovery of *CSF3R* mutations (see above) and their almost invariable association with WHO-defined CNL has allowed its incorporation in the WHO diagnostic criteria (Table 110-5). In practical terms, the presence of a membrane proximal *CSF3R* mutation in a patient with predominantly neutrophilic granulocytosis should be sufficient for the diagnosis of CNL, regardless of the degree of leukocytosis. Unfortunately, several exclusionary criteria still need to be met for diagnosing CNL in the absence of *CSF3R* mutations (Table 110-5).

Treatment Current treatment in CNL is largely palliative and suboptimal in its efficacy. Several drugs alone or in combination have been tried, and none have shown remarkable efficacy. As such, allogeneic hematopoietic stem cell transplant (AHSCT) is reasonable to consider in the presence of symptomatic disease, especially in younger patients. Otherwise, cytoreductive therapy with hydroxyurea is probably as good as anything, and a more intensive combination chemotherapy may not have additional value. However, response to hydroxyurea therapy is often transient, and some have successfully used interferon α as an alternative drug. Response to treatment with ruxolitinib (a JAK1 and JAK2 inhibitor) has been reported in several case reports, but as is the case with hydroxyurea treatment, the response was often incomplete and temporary. In a recently reported phase 2 study of ruxolitinib in 44 patients with CNL or aCML, 21 patients had CNL (76% harbored *CSF3R* mutations), of whom only 4 (20%) experienced complete or partial response according to conventional response criteria.

TABLE 110-4 World Health Organization Classification of Myeloid Malignancies

1. Acute myeloid leukemia (AML) and related precursor neoplasms
2. Myeloproliferative neoplasms (MPN)
 - 2.1. Chronic myeloid leukemia (CML), *BCR-ABL1* positive
 - 2.2. JAK2 mutation-enriched MPN
 - 2.2.1. Polycythemia vera
 - 2.2.2. Primary myelofibrosis
 - 2.2.3. Essential thrombocythemia
 - 2.3. Chronic neutrophilic leukemia (CNL)
 - 2.4. Chronic eosinophilic leukemia, not otherwise specified (CEL-NOS)
 - 2.5. Myeloproliferative neoplasm, unclassifiable (MPN-U)
3. Myelodysplastic syndromes (MDS)
 - 3.1. MDS with single lineage dysplasia
 - 3.2. MDS with ring sideroblasts (MDS-RS)
 - 3.3. MDS with multilineage dysplasia
 - 3.4. MDS with excess blasts
 - 3.5. MDS with isolated del(5q)
 - 3.6. MDS, unclassifiable (MDS-U)
 - 3.7. Provisional entity: Refractory cytopenia of childhood
4. MDS/MPN overlap
 - 4.1. Chronic myelomonocytic leukemia (CMML)
 - 4.2. Atypical chronic myeloid leukemia (aCML), *BCR-ABL1* negative
 - 4.3. Juvenile myelomonocytic leukemia (JMML)
 - 4.4. MDS/MPN with ring sideroblasts and thrombocytosis (MDS/MPN-RS-T)
 - 4.5. MDS/MPN, unclassifiable (MDS/MPN-U)
5. Mastocytosis
6. Myeloid/lymphoid neoplasms with eosinophilia and rearrangement of *PDGFRA*, *PDGFRB*, or *FGFR1* or with *PCM1-JAK2*
 - 6.1. Myeloid/lymphoid neoplasms with *PDGFRA* rearrangement
 - 6.2. Myeloid/lymphoid neoplasms with *PDGFRB* rearrangement
 - 6.3. Myeloid/lymphoid neoplasms with *FGFR1* rearrangement
 - 6.4. Provisional entity: Myeloid/lymphoid neoplasms with *PCM1-JAK2* translocation
7. Myeloid neoplasms with germline predisposition

TABLE 110-5 2016 World Health Organization (WHO) Diagnostic Criteria for Chronic Neutrophilic Leukemia (CNL), Atypical Chronic Myeloid Leukemia, BCR-ABL1-Negative (aCML), and Chronic Myelomonocytic Leukemia (CMML)

VARIABLES	CNL	aCML	CMML
PB leukocyte count	$\geq 25 \times 10^9/L$	Granulocytosis	
PB segmented neutrophils/bands	$\geq 80\%$		
PB immature granulocytes ^a	<10%	$\geq 10\%$	
PB blast count	<1%	<20%	<20%
PB monocyte count	$<1 \times 10^9/L$	No or minimal monocytosis	$\geq 1 \times 10^9/L$ Persistent and lasting for at least 3 months
Dysgranulopoiesis	No	Yes	
PB basophil percentage		<2%	
PB monocyte percentage		<10%	$\geq 10\%$
BM	Hypercellular \uparrow Neutrophils, number and % <5% blasts Normal neutrophilic maturation	Hypercellular \uparrow Granulocyte proliferation Granulocytic dysplasia \pm erythroid/megakaryocyte Dysplasia <20% blasts	Dysplasia in ≥ 1 myeloid lineages or Clonal cytogenetic/molecular abnormality <20% blasts or promonocytes
BCR-ABL1	No	No	No
PDGFRA, PDGFRB, FGFR1, or PCM1-JAK2 rearrangement	No	No	No
CSF3R T618I or other activating CSF3R mutation or persistent neutrophilia, splenomegaly, and no identifiable cause of reactive neutrophilia	Yes		
PB and BM blasts/promonocytes		<20%	<20%
Evidence for other MPN: CML, PV, ET, or PMF	No	No	No
Evidence for reactive leukocytosis ^b or monocytosis	No		No

^aImmature granulocytes include myeloblasts, promyelocytes, myelocytes, and metamyelocytes. ^bCauses of reactive neutrophilia include plasma cell neoplasms, solid tumor, infections, and inflammatory processes.

Abbreviations: BM, bone marrow; CML, chronic myeloid leukemia; ET, essential thrombocythemia; MPN, myeloproliferative neoplasms; PB, peripheral blood; PMF, primary myelofibrosis; PV, polycythemia vera.

ATYPICAL CHRONIC MYELOID LEUKEMIA

Atypical chronic myeloid leukemia, *BCR-ABL1* negative (aCML) is formally classified under the MDS/MPN category of myeloid malignancies and is characterized by left-shifted granulocytosis and dysgranulopoiesis. The differential diagnosis of aCML includes CML, which is distinguished by the presence of *BCR-ABL1*; CNL, which is distinguished by the absence of dysgranulopoiesis and presence of *CSF3R* mutations; and CMML, which is distinguished by the presence of monocytosis (absolute monocyte count $\geq 1 \times 10^9/L$). The WHO diagnostic criteria for aCML are listed in Table 110-5 and include granulocytosis; dysgranulopoiesis; $\geq 10\%$ immature granulocytes; <20% PB or BM myeloblasts; <10% PB monocytes; <2% basophils; absence of otherwise specific mutations such as *BCR-ABL1*, *PDGFRA*, *PDGFRB*, *FGFR1*, or *PCM1-JAK2*; and not meeting WHO criteria for CML, primary myelofibrosis (PMF), polycythemia vera (PV), or essential thrombocythemia (ET). The BM in aCML is hypercellular with granulocyte proliferation and dysplasia with or without erythroid or megakaryocytic dysplasia.

The molecular pathogenesis of aCML is incompletely understood; about a fourth of patients express *SETBP1* mutations, which are, however, also found in several other myeloid malignancies, including CNL and CMML. *SETBP1* mutations in aCML are prognostically detrimental and mostly located between codons 858 and 871; similar mutations are seen with Schinzel-Giedion syndrome (a congenital disease with severe developmental delay and various physical stigmata including midface retraction, large forehead, and macroglossia). More recently, a somatic missense mutation in ethanolamine kinase 1 (*ETNK1* N244S) was described in 9% of patients with aCML but was also seen in 14% of patients with CMML, 6% of patients with mastocytosis (especially in association with eosinophilia), and rarely in other MPNs.

In a series of 55 patients with WHO-defined aCML, median age at diagnosis was 62 years with female preponderance (57%); splenomegaly was reported in 54% of the patients, red cell transfusion requirement in 65%, abnormal karyotype in 20% (20q- and trisomy 8 being the most frequent), and leukemic transformation in 40%. Median survival was 25 months. Outcome was worse in patients with marked leukocytosis, transfusion requirement, and increased immature cells in the PB. In a more recent Mayo Clinic study of 25 molecularly annotated and strictly WHO-defined aCML patients, median age was 70 years and 84% were male. Cytogenetic abnormalities were seen in 36% and gene mutations in 100%. Mutational frequencies were as follows: *ASXL1* 28%, *TET2* 16%, *NRAS* 16%, *SETBP1* 12%, *RUNX1* 12%, *ETNK1* 8%, and *PTPN11* 4%. Median survival was 10.8 months, and at last follow-up (median 11 months), 17 deaths (68%) and 2 leukemic transformations (8%) were documented. In multivariable analysis, advanced age, low hemoglobin, and *TET2* mutations were shown to carry independent prognostic significance; other mutations, including *ASXL1* and *SETBP1*, lacked prognostic significance. Conventional chemotherapy is largely ineffective in the treatment of aCML. Similarly, treatment response to the JAK1/2 inhibitor ruxolitinib has not been impressive. However, a favorable experience with autologous stem cell transplantation (ASCT) was reported in nine patients; after a median follow-up of 55 months, the majority of the patients remained in complete remission.

CHRONIC MYELOMONOCYTIC LEUKEMIA

CMML is classified under the WHO category of MDS/MPN and is defined by an absolute monocyte count (AMC) of $\geq 1 \times 10^9/L$ in the PB and accounting for $\geq 10\%$ of the leukocyte count. Median age at diagnosis ranges from 65 to 75 years, and there is a 2:1 male predominance.

Clinical presentation is variable and depends on whether the disease presents with MDS-like or MPN-like phenotype; the former is associated with cytopenias and the latter with splenomegaly and features of myeloproliferation such as fatigue, night sweats, weight loss, and cachexia. About 20% of patients with CMML experience serositis involving the joints (arthritis), pericardium (pericarditis and pericardial effusion), pleura (pleural effusion), or peritoneum (ascites).

Pathogenesis Almost all patients with CMML harbor somatic mutations involving epigenetic regulator genes (e.g., ASXL1, TET2), spliceosome pathway genes (e.g., SRSF2), DNA damage response genes (e.g., TP53), and tyrosine kinases/transcription factors (e.g., KRAS, NRAS, CBL, and RUNX1). However, none of these mutations are specific to CMML, and their precise pathogenetic contribution is unclear. Clonal cytogenetic abnormalities are seen in about a third of patients with CMML and include trisomy 8 and abnormalities of chromosome 7. More recent studies have demonstrated the presence of BM dendritic cell aggregates suggesting systemic immune dysregulation and distinct phenotypic features of monocytes in CMML.

Diagnosis Reactive monocytosis is uncommon but has been reported in association with certain infections and inflammatory conditions. Clonal (i.e., neoplastic) monocytosis defines CMML but is also seen with JMML and AML with monocytic differentiation. The WHO diagnostic criteria for CMML are listed in Table 110-5 and include persistent PB monocyte count of $\geq 1 \times 10^9/L$ with monocyte percentage of $\geq 10\%$; absence of *BCR-ABL1*, *PDGFRA*, *PDGFRB*, *FGFR1*, or *PCM1-JAK2* rearrangements; not meeting WHO criteria for CML, PV, ET, or PMF; <20% blasts and promonocytes in the PB and BM; and dysplasia involving one or more myeloid lineages or, in the absence of dysplasia, presence of an acquired clonal cytogenetic or molecular genetic abnormality or nonreactive monocytosis lasting for at least 3 months.

The BM in CMML is hypercellular with granulocytic and monocytic proliferation. Dysplasia is often present and may involve one, two, or all myeloid lineages. On immunophenotyping, the abnormal cells often express myelomonocytic antigens such as CD13 and CD33, with variable expression of CD14, CD68, CD64, and CD163. Monocytic-derived cells are almost always positive for the cytochemical non-specific esterases (e.g., butyrate esterase), while normal granulocytic precursors are positive for lysozyme and chloroacetate esterase. In CMML, it is common to have a hybrid cytochemical staining pattern with cells expressing both chloroacetate and butyrate esterases simultaneously (dual esterase staining). Monocytosis can be associated with reactive as well as other myeloid neoplasms. Based on flow cytometric expression of CD14/CD16, monocytes can be classified into classical MO1 (CD14+/CD16-), intermediate MO2 (CD14+/CD16+), and nonclassical MO3 (CD14-/CD16+) fractions, with MO1 constituting the major monocyte population (85%) in healthy conditions. Recent studies have suggested characteristic increase in classical monocytes in CMML patients, distinguishing them from other causes of reactive and clonal monocytosis.

Prognosis A recent meta-analysis showed median survival of 1.5 years in CMML. Numerous prognostic systems have attempted to better define and stratify the natural history of CMML. One of these, the Mayo prognostic model, assigns one point each to the following four independent prognostic variables: $AMC > 10 \times 10^9/L$, presence of circulating immature cells, hemoglobin $< 10 \text{ g/dL}$, and platelet count $< 100,000/\text{mL}$. This model stratified patients into three risk groups: low (0 points), intermediate (1 point), and high (≥ 2 points), translating to median survival of 32, 18, and 10 months, respectively. Another prognostic model referred to as the CMML-specific prognostic scoring system (CPSS) identified four variables as being prognostic for both overall survival and leukemia-free survival: French-American-British (FAB) and WHO CMML subtypes, red blood cell transfusion dependency, and the Spanish cytogenetic risk stratification system. A French study incorporated ASXL1 mutational status in 312 CMML patients; in a multivariable model, independent predictors of poor survival were WBC $> 15 \times 10^9/L$ (3 points), ASXL1 mutations (2 points), age > 65 years (2 points), platelet count $< 100,000/\text{mL}$ (2 points), and

hemoglobin $< 10 \text{ g/dL}$ in females and $< 11 \text{ g/dL}$ in males (2 points). This model stratified patients into three groups: low (0–4 points), intermediate (5–7 points), and high risk (8–12 points), with median survivals of not reached and 38.5 and 14.4 months, respectively. More recent studies have highlighted the adverse prognostic effect of ASXL1 and DNMT3A mutations in CMML. To further clarify the prognostic relevance of ASXL1 mutations, an international collaborative cohort of 466 CMML patients was analyzed. In multivariable analysis, ASXL1 mutations, $AMC > 10 \times 10^9/L$, hemoglobin $< 10 \text{ g/dL}$, platelets $< 100 \times 10^9/L$, and circulating immature myeloid cells were independently predictive of shortened overall survival. More recently, the aforementioned CPSS model was updated to include molecular abnormalities including ASXL1, RUNX1, NRAS, and SETBP1 mutations (CPSS-Mol). In a report of 171 patients with blast phase CMML (median age 71 years), treatment included best supportive care in 25%, hypomethylating agent therapy in 10%, AML-like induction chemotherapy in 38%, AML-like induction chemotherapy followed by AHSCT in 15%, upfront AHSCT in 2%, and clinical trials in 11%. After a median follow-up of 4.4 months, 141 deaths (82%) were recorded. Median overall survival was 6 months with 1-, 3-, and 5-year survival rates of 25%, 9%, and 6%, respectively.

Treatment Current treatment in CMML consists of hydroxyurea and supportive care, including red cell transfusions and use of erythropoiesis-stimulating agents (ESAs). The value of hydroxyurea was reinforced by a randomized trial against oral etoposide. No other single or combination chemotherapy has been shown to be superior to hydroxyurea. AHSCT is a viable treatment option for transplant-eligible patients with poor prognostic features. Given the MDS/MPN overlap phenotype and the presence of MDS-like genetic/methylation abnormalities in CMML, hypomethylating agents such as 5-azacitidine and decitabine have been used with limited efficacy; in a recent study using decitabine in CMML, overall response rate was 48% with 17% complete remissions and median survival of 17 months. The experience with 5-azacytidine was somewhat similar. In a recent Mayo Clinic report, among 406 consecutive CMML patients (age ≤ 75 years at diagnosis) seen between January 1990 and December 2018, 70 (17%) underwent AHSCT (median age 58 years) including 46 (66%) in chronic phase and 24 (34%) in blast phase. At a median follow-up of 70 months, there were 22 deaths (31%) in the chronic phase transplant group, 11 (24%) from disease relapse and 9 (20%) from nonrelapse mortality. Posttransplant median survival was 67 months in the chronic phase and 16 months in the blast phase ($p < .01$) transplant groups; 5-year survival rates were 51% and 19%, respectively.

■ JUVENILE MYELOMONOCYTIC LEUKEMIA

JMML is primarily a disease of early childhood and is included, along with CMML, in the MDS/MPN WHO category. Both CMML and JMML feature leukocytosis, monocytosis, and hepatosplenomegaly. Additional characteristic features in JMML include thrombocytopenia and elevated fetal hemoglobin. Myeloid progenitors in JMML display granulocyte-macrophage colony-stimulating factor (GM-CSF) hypersensitivity that has been attributed to dysregulated RAS/MAPK signaling. The latter is believed to result from mutually exclusive mutations involving *RAS*, *PTPN11*, and *NF1*. A third of patients with JMML that is not associated with Noonan syndrome carry *PTPN11* mutations, whereas the incidence of *NF1* in patients without neurofibromatosis type 1 (*NF1*) and *RAS* mutations is ~15% each. In general, ~85% of JMML cases have one of the classical RAS pathway mutations (*PTPN11*, *NRAS*, *KRAS*, *NF1*, or *CBL*); in addition, a myriad of other mutations, such as ASXL1, RUNX1, SETBP1, JAK3, and CUX1, have recently been reported. Drug therapy is relatively ineffective in JMML, and the treatment of choice is AHSCT, which results in a 5-year survival of approximately 50%.

The 2016 revised WHO diagnostic criteria for JMML require the presence of PB monocyte count $\geq 1 \times 10^9/L$, <20% blasts in blood or BM, splenomegaly, and absence of *BCR-ABL1*. Diagnosis also requires the presence of one of the following: somatic mutation of *PTPN11*,

KRAS, or NRAS; clinical diagnosis of NF1 or *NF1* mutation; germline mutation of *CBL*; and loss of heterozygosity. Diagnosis of JMML can still be considered without the aforementioned genetic features in the presence of monosomy 7 or any other cytogenetic abnormality or in the presence of two of the following: increased hemoglobin F, presence of myeloid or erythroid precursors in the PB, GM-CSF hypersensitivity in colony assay, and hyperphosphorylation of STAT5.

MDS/MPN, UNCLASSIFIABLE (MDS/MPN-U)

The WHO classifies patients with morphologic and laboratory features that resemble both MDS and MPN as “MDS/MPN overlap.” This category includes CMML, aCML, and JMML, which have been discussed above. In addition, MDS/MPN includes a fourth category referred to as MDS/MPN, unclassifiable (MDS/MPN-U). Diagnosis of MDS/MPN-U requires the presence of both MDS and MPN features that are not adequate to classify patients as CMML, aCML, or JMML. MDS/MPN also includes the provisional category of refractory anemia with ring sideroblasts and thrombocytosis (RARS-T); the 2016 revision of the WHO classification document has changed the term RARS-T to MDS/MPN-RS-T.

In a representative study of 85 patients with MDS/MPN-U, median age was 70 years and 72% were male. Splenomegaly at presentation was present in 33%, thrombocytosis in 13%, leukocytosis in 18%, *JAK2* mutations in 30%, and abnormal karyotype in 51%; the most frequent cytogenetic abnormality was trisomy 8. Median survival was 12.4 months and favorably affected by thrombocytosis. Treatment with hypomethylating agents, immunomodulators, or AHSCT did not appear to favorably affect survival.

MDS/MPN WITH RING SIDEROBLASTS AND THROMBOCYTOSIS (MDS/MPN-RS-T)

MDS/MPN-RS-T is classified in the MDS/MPN category because it shares dysplastic features with MDS-RS and myeloproliferative features with ET. The 2016 revised WHO diagnostic criteria for MDS/MPN-RS-T includes anemia associated with erythroid lineage dysplasia, presence of $\geq 15\%$ ring sideroblasts, blast count of $< 5\%$ in BM and $< 1\%$ in the PB, platelet count of $\geq 450 \times 10^9/L$, and absence of *BCR-ABL1*, *PDGFRA*, *PDGFRB*, *FGFR1*, or *PCM1-JAK2* mutations or t(3;3) (q21;q26), inv(3)(q21q26), or del(5q). These new diagnostic criteria also require the absence of history of MPN, MDS, or other type of MDS/MPN and also either the presence of *SF3B1* mutation or absence of exposure to cytotoxic or other treatment that could be blamed for the morphologic abnormalities.

In a recent study, 111 patients with MDS/MPN-RS-T were compared with 33 patients with RARS. The frequency of *SF3B1* mutations in MDS/MPN-RS-T-T (87%) was similar to that in MDS-RS (85%). *JAK2* V617F mutation was detected in 49% of MDS/MPN-RS-T patients (including 48% of those mutated for *SF3B1*) but none of those with MDS-RS. In MDS/MPN-RS-T, *SF3B1* mutations were more frequent in females (95%) than in males (77%), and mean ring sideroblast counts were higher in *SF3B1*-mutated patients. Median overall survival was 6.9 years in *SF3B1* mutated cases versus 3.3 years in unmutated cases. Six-year survival was 67% in *JAK2* mutated cases versus 32% in unmutated cases. Multivariable analysis identified younger age and *JAK2* and *SF3B1* mutations as favorable factors. Predictors of poor survival in MDS/MPN-RS-T include anemia, abnormal karyotype, and presence of *ASXL1* or *SETBP1* mutations. Interestingly, the presence of *SF3B1* mutations in MDS/MPN-RS-T was recently shown to be associated with increased risk of thrombosis. Several case reports have suggested that treatment with lenalidomide might induce red cell transfusion independency and complete remissions in MDS/MPN-RS-T. Most recently, luspatercept, a recombinant fusion protein that binds transforming growth factor β superfamily ligands to reduce SMAD signaling, has also been shown to benefit some patients with MDS-RS-T; in a recently published phase 3 trial involving 229 patients with transfusion-dependent very-low- to intermediate-risk MDS-RS-T, transfusion independence for ≥ 8 weeks was achieved in 38% of the patients receiving luspatercept versus 13% of patients in the placebo group ($p < 0.1$).

■ MYELOPROLIFERATIVE NEOPLASM, UNCLASSIFIABLE (MPN-U)

The category of MPN-U includes MPN-like neoplasms that cannot be clearly classified as one of the other seven subcategories of MPN (Table 110-4). Examples include patients presenting with unusual thrombosis or unexplained organomegaly with normal blood counts but found to carry MPN-characteristic mutations such as *JAK2* and *CALR* or display BM morphology that is consistent with MPN. It is possible that some cases of MPN-U represent earlier disease stages in PV or ET, which however fail to meet the threshold hemoglobin levels or platelet counts that are required per WHO diagnostic criteria. Specific treatment interventions might not be necessary in asymptomatic patients with MPN-U, whereas patients with arterial thrombotic complications might require cytoreductive and aspirin therapy and those with venous thrombosis might require systemic anticoagulation.

■ MYELOID NEOPLASMS WITH GERMLINE PREDISPOSITION

The 2016 WHO revision on the classification of myeloid neoplasms added a section referred to as “myeloid neoplasms with germline predisposition” and that includes cases of AML, MDS, and MDS/MPN that arise in the setting of a germline predisposition mutation, such as *CEBPA*, *DDX41*, *RUNX1*, *ANKRD26*, *ETV6*, or *GATA2*. This particular category of diseases also includes myeloid neoplasms that arise in the background of BM failure syndromes, Down syndrome, Noonan syndrome, neurofibromatosis, and telomopathies.

■ TRANSIENT MYELOPROLIFERATIVE DISORDER (TMD)

TMD, also referred to as transient abnormal myelopoiesis (TAM), constitutes an often but not always transient phenomenon of abnormal megakaryoblast proliferation, which occurs in $\sim 10\%$ of infants with Down syndrome. TMD is usually recognized at birth and either undergoes spontaneous regression (75% of cases) or progresses to acute megakaryoblastic leukemia (AMKL) (25% of cases). Almost all patients with TMD and TMD-derived AMKL display somatic *GATA1* mutations. TMD-associated *GATA1* mutations constitute exon 2 insertions, deletions, or missense mutations, affecting the N-terminal transactivation domain of GATA-1, and result in loss of full-length (50-kD) GATA-1 and its replacement with a shorter isoform (40-kD) that retains friend of GATA-1 (FOG-1) binding. In contrast, inherited forms of exon 2 *GATA1* mutations produce a phenotype with anemia, whereas exon 4 mutations that affect the N-terminal, FOG-1-interactive domain produce familial dyserythropoietic anemia with thrombocytopenia or X-linked macrothrombocytopenia.

■ PRIMARY EOSINOPHILIA

Eosinophilia refers to a PB absolute eosinophil count (AEC) that is above the upper normal limit of the reference range. The term *hyper-eosinophilia* is used when the AEC is $> 1500 \times 10^9/L$. Eosinophilia is operationally classified into secondary (nonneoplastic proliferation of eosinophils) and primary (proliferation of eosinophils that is either neoplastic or otherwise unexplained). Secondary eosinophilia is by far the most frequent cause of eosinophilia and is often associated with infections, especially those related to tissue-invasive helminths, allergic/vasculitic diseases, drugs, and metastatic cancer. Primary eosinophilia is the focus of this chapter and is considered when a cause for secondary eosinophilia is not readily apparent.

Primary eosinophilia is classified as clonal or idiopathic. Diagnosis of clonal eosinophilia requires morphologic, cytogenetic, or molecular evidence of a myeloid neoplasm. Idiopathic eosinophilia is considered when both secondary and clonal eosinophilias have been ruled out as a possibility. HES is a subcategory of idiopathic eosinophilia with persistent AEC of $\geq 1.5 \times 10^9/L$ and associated with eosinophil-mediated organ damage (Table 110-6). An HES-like disorder that is associated with clonal or phenotypically abnormal T cells is referred to as lymphocytic variant hypereosinophilia (Table 110-6).

Clonal Eosinophilia Examples of clonal eosinophilia include eosinophilia associated with AML, MDS, CML, mastocytosis, and MDS/

TABLE 110-6 Primary Eosinophilia Classification

VARIABLES	EOSINOPHILIA ASSOCIATED WITH PDGFRA, PDGFRB, FGFR1, OR PCM1-JAK2 ABNORMALITY	CHRONIC EOSINOPHILIA NOT OTHERWISE SPECIFIED (CEL-NOS)	LYMPHOCYTIC VARIANT HYPEREOSINOPHILIA	HYPEREOSINOPHILIC SYNDROME
Absolute eosinophil count	$>600 \times 10^9/L$	$>1500 \times 10^9/L$	$>1500 \times 10^9/L$	$>1500 \times 10^9/L$
Peripheral blood blasts >2%	Yes or no	Yes or no	No	No
Bone marrow blasts >5%	Yes or no	Yes or No	No	No
Abnormal karyotype	Yes or no	Yes or no	No	No
PDGFRA, PDGFRB, FGFR1, or PCM1-JAK2 abnormality	Yes	No	No	No
BCR-ABL1	No	No	No	No
Abnormal T lymphocyte phenotype or clonal T-cell clones	No	No	Yes	No
Eosinophil-mediated tissue damage	Yes or no	Yes or no	Yes or no	Yes

MPN overlap. Myeloid neoplasm-associated eosinophilia also includes the WHO MPN subcategory of chronic eosinophilic leukemia, not otherwise specified (CEL-NOS) and the WHO myeloid malignancy subcategory referred to as myeloid/lymphoid neoplasms with eosinophilia and rearrangement of platelet-derived growth factor receptor (*PDGFR*) α/β or fibroblast growth factor receptor 1 (*FGFR1*) or with *PCM1-JAK2* (Table 110-4).

The diagnostic workup for clonal eosinophilia that is not associated with morphologically overt myeloid malignancy should start with PB mutation screening for *FIP1L1-PDGFR*A and *PDGFRB* mutations using fluorescence in situ hybridization (FISH) or reverse transcription polymerase chain reaction. This is crucial since such eosinophilia is easily treated with imatinib. If mutation screening is negative, a BM examination with cytogenetic studies is indicated. In this regard, one must first pay attention to the presence or absence of 5q33, 4q12, 8p11.2, or t(8;9)(p22;p24.1) translocations, which, if present, would suggest *PDGFRB*-, *PDGFR*A-, or *FGFR1*-rearranged or *PCM1-JAK2*-associated clonal eosinophilia, respectively. The presence of 5q33 or 4q12 translocations predicts favorable response to treatment with imatinib mesylate and presence of t(8;9)(p22;p24.1) predicts a transient response to ruxolitinib, whereas 8p11.2 translocations are associated with aggressive myeloid malignancies that are refractory to current drug therapy.

Chronic Eosinophilic Leukemia, Not Otherwise Specified (CEL-NOS) CEL-NOS is a subset of clonal eosinophilia that is neither molecularly defined nor classified as an alternative clinicopathologically assigned myeloid malignancy. We prefer to use the term strictly in patients with an HES phenotype who also display either a clonal cytogenetic/molecular abnormality or excess blasts in the BM or PB. The WHO defines CEL-NOS as the presence of $\geq 1.5 \times 10^9/L$ AEC that is accompanied by either the presence of myeloblast excess (either >2% in the PB or 5–19% in the BM) or evidence of myeloid clonality.

In a recent Mayo Clinic survey of 1416 patients with PB eosinophilia evaluated between 2008 and 2019, 17 patients (1.2%) fulfilled the WHO 2016 criteria for CEL-NOS. Median age was 63 years (range 25–92 years) with the vast majority of patients (88%) presenting with systemic symptoms. Organ involvement was a prominent feature, and involved organs included spleen, cardiac and pulmonary organs, and distal esophagus. Laboratory abnormalities included anemia, leukocytosis, and eosinophilia (median eosinophil count of $6.4 \times 10^9/L$; range 2.0–53.1). The most common BM abnormalities included abnormal eosinophils, abnormal and increased megakaryocytes, and fibrosis (18%). Cytogenetic abnormalities occurred in 88% of patients and included trisomy 8, complex karyotype, 13q-, 20q-, and chromosome 1 abnormalities. All seven patients with next-generation sequencing studies harbored one or more mutations, including *ASXL1* (43%) and *IDH1* (29%). Half of patients treated with hydroxyurea-based regimens responded with a persistent decline in eosinophil count for a median duration of 18 months. One-third of patients treated with prednisone responded, with a median duration of response of 13 months. Three

patients were treated with imatinib, of whom two had normalization of eosinophil count. At a median follow-up of 13 months, nine patients had died including three patients who underwent leukemic transformation.

PDGFR Mutated Eosinophilia Both platelet-derived growth factor receptors α (*PDGFRA* located on chromosome 4q12) and β (*PDGFRB* located on chromosome 5q31-q32) are involved in MPN-relevant activating mutations. Clinical phenotype in both instances includes prominent blood eosinophilia and excellent response to imatinib therapy. In regard to *PDGFRA* mutations, the most popular is *FIP1L1-PDGFR*A, a karyotypically occult del(4)(q12), that was described in 2003 as an imatinib-sensitive activating mutation. Functional studies have demonstrated transforming properties in cell lines and the induction of MPN in mice. Cloning of the *FIP1L1-PDGFR*A fusion gene identified a novel molecular mechanism for generating this constitutively active fusion tyrosine kinase, wherein a ~800-kb interstitial deletion within 4q12 fuses the 5' portion of *FIP1L1* to the 3' portion of *PDGFRA*. *FIP1L1-PDGFR*A occurs in a very small subset of patients who present with the phenotypic features of either systemic mastocytosis (SM) or HES, but the presence of the mutation reliably predicts complete hematologic and molecular response to imatinib therapy.

In a recent retrospective survey of 151 patients with *FIP1L1-PDGFR*A-associated eosinophilia (143 males; mean age at diagnosis 49 years), organopathy involved the spleen (44%), skin (32%), lungs (30%), heart (19%), and central nervous system (9%); none of 31 patients initially treated with corticosteroids achieved complete hematologic remission, whereas all 148 patients treated with imatinib achieved complete hematologic responses and also molecular responses, when evaluated. Treatment discontinuation was documented in 46 patients followed by a 57% relapse rate; the 1-, 5-, and 10-year overall survival rates in imatinib-treated patients were 99%, 95%, and 84%, respectively. Other studies have confirmed the possibility of treatment-free remissions in some patients after imatinib discontinuation. Infrequent occurrence of *FIP1L1-PDGFR*A mutated acute myeloid leukemia associated with eosinophilia has also been shown to respond to low-dose imatinib therapy (100 mg/d).

The association between eosinophilic myeloid malignancies and *PDGFRB* rearrangement was first characterized and published in 1994 where fusion of the tyrosine kinase encoding region of *PDGFRB* to the *ets*-like gene *ETV6* (*ETV6-PDGFRB*, t[5;12][q33;p13]) was demonstrated. The fusion protein was transforming to cell lines and resulted in constitutive activation of *PDGFRB* signaling. Since then, several other *PDGFRB* fusion transcripts with similar disease phenotypes have been described, and cell line transformation and MPD induction in mice have been demonstrated. Imatinib therapy was shown to be effective when employed.

FGFR1 Mutated Eosinophilia The 8p11 myeloproliferative syndrome (EMS) (also known as human stem cell leukemic/lymphoma syndrome) constitutes a clinical phenotype with features of both lymphoma and eosinophilic MPN and is characterized by a fusion

mutation that involves the gene for fibroblast growth factor receptor-1 (*FGFR1*), which is located on chromosome 8p11. In EMS, both myeloid and lymphoid lineage cells exhibit the 8p11 translocation, thus demonstrating the stem cell origin of the disease. The disease features several 8p11-linked chromosome translocations, and some of the corresponding fusion *FGFR1* mutants have been shown to transform cell lines and induce EMS- or CML-like disease in mice depending on the specific *FGFR1* partner gene (*ZNF198* or *BCR*, respectively). Consistent with this laboratory observation, some patients with *BCR-FGFR1* mutation manifest a more indolent CML-like disease. The mechanism of *FGFR1* activation in EMS is similar to that seen with *PDGFRB*-associated MPD; the tyrosine kinase domain of *FGFR1* is juxtaposed to a dimerization domain from the partner gene. EMS is aggressive and requires combination chemotherapy followed by ASCT.

PCM1-JAK2-Associated Myeloid/Lymphoid Neoplasm with Eosinophilia The 2016 revised WHO document includes a provisional entity under myeloid/lymphoid neoplasms with eosinophilia referred to as “myeloid/lymphoid neoplasms with *PCM1-JAK2*.² The entity is characterized by the t(8;9)(p22;p24.1) cytogenetic abnormality and a phenotype that displays marked male predominance, hepatosplenomegaly, eosinophilia, and morphologic features similar to MPN, MDS, or MDS/MPN. Current drug therapy for *PCM1-JAK2*-associated disease is suboptimal, although some affected patients have displayed transient responses to ruxolitinib therapy.

Hypereosinophilic Syndrome Blood eosinophilia that is neither secondary nor clonal is operationally labeled as being “idiopathic.” HES is a subcategory of idiopathic eosinophilia with persistent increase of the AEC to $\geq 1.5 \times 10^9/L$ and presence of eosinophil-mediated organ damage, including cardiomyopathy, gastroenteritis, cutaneous lesions, sinusitis, pneumonitis, neuritis, and vasculitis. In addition, some patients manifest thromboembolic complications, hepatosplenomegaly, and either cytopenia or cytosis.

BM histologic and cytogenetic/molecular studies should be examined before a working diagnosis of HES is made. Additional blood studies that are currently recommended during the evaluation of HES include serum tryptase (an increased level suggests mastocytosis and warrants molecular studies to detect *FIP1L1-PDGFRα*), T-cell immunophenotyping, and T-cell receptor antigen gene rearrangement analysis (a positive test suggests an underlying clonal or phenotypically abnormal T-cell disorder). In addition, initial evaluation in HES should include echocardiogram and measurement of serum troponin levels to screen for myocardial involvement by the disease.

Initial evaluation of the patient with eosinophilia should include tests that facilitate assessment of target organ damage: complete blood count, chest x-ray, echocardiogram, and serum troponin level. Increased level of serum cardiac troponin has been shown to correlate with the presence of cardiomyopathy in HES. Typical echocardiographic findings in HES include ventricular apical thrombus, posterior mitral leaflet or tricuspid valve abnormality, endocardial thickening, dilated left ventricle, and pericardial effusion.

In a recent Mayo Clinic study of 98 consecutive patients with idiopathic eosinophilia, including HES, median age was 53 years (55% males) and overt organ involvement was seen in >80% of the cases, including 54% involving organs other than the skin. The frequencies of cardiac involvement, hepatosplenomegaly, and increased serum tryptase and interleukin 5 (IL-5) levels were 8%, 4%, 24%, and 31%, respectively. The study also revealed that 11% of the affected patients harbored pathogenetic mutations including *TET2*, *ASXL1*, and *KIT*; the presence of such mutations did not appear to influence phenotype, and the number of informative cases was too small to assess prognostic relevance. Instead, the study identified anemia and presence of cardiac involvement or hepatosplenomegaly as risk factors for survival.

Corticosteroids are the cornerstone of therapy in HES. Treatment with oral prednisone is usually started at 1 mg/kg/d and continued for 1–2 weeks before the dose is tapered slowly over the ensuing 2–3 months. If symptoms recur at a prednisone dose level of >10 mg/d, either hydroxyurea or interferon α is used as a steroid-sparing agent. In patients who fail usual therapy as outlined above, mepolizumab

or alemtuzumab might be considered. Mepolizumab is a monoclonal antibody that targets IL-5, which is a well-recognized survival factor for eosinophils. Alemtuzumab targets the CD52 antigen, which has been shown to be expressed by eosinophils but not by neutrophils. In a recently reported, placebo-controlled, phase 3 study, HES patients received subcutaneous mepolizumab (300 mg) every 4 weeks, in addition to their preprotocol therapy, and experienced significantly fewer disease flare ups or treatment discontinuations (28 vs 56% for placebo), without excess adverse events. Mepolizumab was approved by the U.S. Food and Drug Administration (FDA) for use in HES on September 25, 2020. In a smaller phase 2 study, benralizumab (monoclonal antibody targeting the receptor for IL-5; 30 mg given subcutaneously every 4 weeks) was also shown to reduce eosinophil count more efficiently compared to placebo (90 vs 30%).

■ MASTOCYTOSIS

Mast cell disease (MCD) is defined as tissue infiltration by morphologically and immunophenotypically abnormal mast cells. MCD is classified into two broad categories: cutaneous mastocytosis (CM) and systemic mastocytosis (SM). MCD in adults is usually systemic, and the clinical course can be either indolent or aggressive, depending on the respective absence or presence of impaired organ function. Symptoms and signs of MCD include urticaria pigmentosa, mast cell mediator release symptoms (e.g., headache, flushing, lightheadedness, syncope, anaphylaxis, pruritus, urticaria, angioedema, nausea, diarrhea, abdominal cramps), and organ damage (lytic bone lesions, osteoporosis, hepatosplenomegaly, cytopenia). Aggressive SM can be associated with another myeloid malignancy, including MPN, MDS, or MDS/MPN overlap (e.g., CMML), or present as overt mast cell leukemia. In general, life expectancy is near normal in indolent SM but significantly shortened in aggressive SM.

Diagnosis of SM is based on BM examination that shows clusters of morphologically abnormal, spindle-shaped mast cells that are best evaluated by the use of immunohistochemical stains that are specific to mast cells (tryptase, CD117). In addition, mast cell immunophenotyping reveals aberrant CD25 expression by neoplastic mast cells. Other laboratory findings in SM include increased levels of serum tryptase, histamine and urine histamine metabolites, and prostaglandins. SM is associated with *KIT* mutations, usually *KIT* D816V, in the majority of patients. Accordingly, mutation screening for *KIT* D816V is diagnostically useful. However, the ability to detect *KIT* D816V depends on assay sensitivity and mast cell content of the test sample. The 2016 WHO classification of mastocytosis includes (1) CM, (2) SM, and (3) mast cell sarcoma (MCS). SM is further classified into (1) indolent SM (ISM), (2) smoldering SM (SSM), (3) SM with an associated hematologic neoplasm (SM-AHN), (4) aggressive SM (ASM), and (5) mast cell leukemia (MCL).

In a recent Mayo Clinic study of 580 patients (median age 55 years; range 18–88 years) with SM, morphologic subcategories were indolent/smoldering in 291 patients (50%) and advanced in 289 patients (50%), including SM-AHN in 199, ASM in 85, and MCL in 5. Multivariable analysis of clinical variables identified age >60 years, advanced SM, thrombocytopenia <150 × 10⁹/L, anemia below sex-adjusted normal, and increased alkaline phosphatase as independent risk factors for survival. In addition, *ASXL1*, *RUNX1*, and *NRAS* mutations were also independently associated with inferior survival. Combined clinical, cytogenetic, and molecular risk factor analysis confirmed the independent prognostic contribution of adverse mutations, advanced SM, thrombocytopenia, increased alkaline phosphatase, and age >60 years. These data were subsequently used to develop clinical and hybrid clinical-molecular risk models. The clinical risk model uses six readily accessible risk variables including age >60 years, platelet count <150 × 10⁹/L, anemia, hypoalbuminemia, increased alkaline phosphatase, and morphologic classification as advanced SM. Accordingly, median survival times were not reached, 148, 65, 31, 18, and 5 months in the presence of ≤1, 2, 3, 4, 5, and 6 of these risk factors, respectively.

Both ISM and ASM patients might experience mast cell mediator release symptoms, which are usually managed by both H₁ and H₂ histamine receptor blockers as well as cromolyn sodium. In addition,

patients with propensity to vasodilatory shock should wear a medical alert bracelet and carry an Epi-Pen self-injector for self-administration of subcutaneous epinephrine. Urticaria pigmentosa shows variable response to both topical and systemic corticosteroid therapy. Cytoreductive therapy is not recommended for ISM, and instead, such patients are managed with use of H_1 and H_2 blockers, leukotriene antagonists, sodium cromolyn, phototherapy, topical steroids, and osteoporosis prevention with diphosphonates including alendronate and pamidronate. In ASM, either interferon α or cladribine is considered first-line therapy and benefits the majority of patients. Cladribine is administered by 2-h infusion ($5\text{ mg}/\text{m}^2$) daily for 5 days, repeated monthly for 4–6 cycles; expected overall response is ~50%, including major response in ~38%. In contrast, imatinib is ineffective in the treatment of *PDGFR* unmutated SM. A controlled study of patients with ISM or SSM demonstrated marginal value of masitinib (oral tyrosine kinase inhibitor that inhibits KIT and LYN), with a reported cumulative symptomatic response rate of 18.7% versus 7.4% for placebo. Treatment responses were more impressive in another study that used the multikinase inhibitor midostaurin in patients with the more aggressive forms of SM, with 45% of the patients achieving major response. Most recently, equally impressive responses were seen with the use of another kinase inhibitor, avapritinib (specifically targets *KIT* D816V), in both ISM and ASM; however, significant drug-related toxicity, including intracranial bleed, cognitive impairment, and moderate to severe cytopenias, has been observed.

DENDRITIC AND HISTIOCYTIC NEOPLASMS

Dendritic cell (DC) and histiocyte/macrophage neoplasms are extremely rare. DCs are antigen-presenting cells, whereas histiocytes/macrophages are antigen-processing. BM myeloid stem cells (CD34+) give rise to monocyte (CD14+, CD68+, CD11c+, CD1a-) and DC (CD14-, CD11c+/-, CD1a+/-) precursors. Monocyte precursors, in turn, give rise to macrophages (CD14+, CD68+, CD11c+, CD163+, lysozyme+) and interstitial DCs (CD68+, CD1a-). DC precursors give rise to Langerhans cell DCs (Birbeck granules, CD1a+, S100+, langerin+) and plasmacytoid DCs (CD68+, CD123+). Follicular DCs (CD21+, CD23+, CD35+) originate from mesenchymal stem cells. Dendritic and histiocytic neoplasms are operationally classified into macrophage/histiocyte-related and DC-related. The former includes histiocytic sarcoma/malignant histiocytosis, and the latter includes Langerhans cell histiocytosis, Langerhans cell sarcoma, interdigitating DC sarcoma, and follicular DC sarcoma.

Histiocytic Sarcoma/Malignant Histiocytosis Histiocytic sarcoma represents malignant proliferation of mature tissue histiocytes and is often localized. Median age at diagnosis is estimated at 46 years with slight male predilection. Some patients might have history of lymphoma, MDS, or germ cell tumors at time of disease presentation. The three typical disease sites are lymph nodes, skin, and gastrointestinal system. Patients may or may not have systemic symptoms including fever and weight loss, and other symptoms include hepatosplenomegaly, lytic bone lesions, and pancytopenia. Immunophenotype includes presence of histiocytic markers (CD68, lysozyme, CD11c, CD14) and absence of myeloid or lymphoid markers. Prognosis is poor and treatment often ineffective. The term *malignant histiocytosis* (MH) refers to a disseminated disease and systemic symptoms. Lymphoma-like treatment induces complete remissions in some patients, and median survival is estimated at 2 years.

Langerhans Cell Histiocytosis Langerhans cells (LCs) are specialized DCs that reside in mucocutaneous tissue and, upon activation, become specialized for antigen presentation to T cells. LC histiocytosis (LCH; also known as histiocytosis X) represents neoplastic proliferation of LCs (S100+, CD1a+, and Birbeck granules on electron microscopy). LCH incidence is estimated at 5 per million, and the disease typically affects children with a male predilection. Presentation can be either unifocal (eosinophilic granuloma) or multifocal. The former usually affects bones and less frequently lymph nodes, skin, and lung, whereas the latter is more disseminated. Unifocal disease often affects older children and adults, whereas multisystem disease affects infants.

LCH of the lung in adults is characterized by bilateral nodules. Prognosis depends on organs involved. Only 10% of patients progress from unifocal to multiorgan disease. LCH of the lung might improve upon cessation of smoking. Approximately 55% of patients with LCH harbor *BRAF* V600E gain-of-function mutations, which indicates high-risk disease and resistance to first-line therapy; however, responses to targeted therapy with vemurafenib have been reported. Other forms of treatment for LCH include combination chemotherapy and MEK inhibitors in *BRAF* wild-type disease with other MAPK pathway mutations. Unfortunately, such targeted therapy has not secured long-lasting, treatment-free remissions.

Langerhans Cell Sarcoma Langerhans cell sarcoma (LCS) also represents neoplastic proliferation of LCs with overtly malignant morphology. The disease can present de novo or progress from antecedent LCH. There is a female predilection, and median age at diagnosis is estimated at 41 years. Immunophenotype is similar to that seen in LCH, and liver, spleen, lung, and bone are the usual sites of disease. Prognosis is poor and treatment generally ineffective.

Interdigitating Dendritic Cell Sarcoma Interdigitating DC sarcoma (IDCS), also known as reticulum cell sarcoma, represents neoplastic proliferation of IDCs. The disease is extremely rare and affects elderly adults with no sex predilection. Typical presentation is asymptomatic solitary lymphadenopathy. Immunophenotype includes S100 positivity and negativity for vimentin and CD1a. Prognosis ranges from benign local disease to widespread lethal disease.

Follicular Dendritic Cell Sarcoma FDCs reside in B-cell follicles and present antigen to B cells. FDC neoplasms (FDCNs) are usually localized and often affect adults. FDCN might be associated with Castleman's disease in 10–20% of cases, and increased incidence in schizophrenia has been reported. Cervical lymph nodes are the most frequent site of involvement in FDCN, and other sites include maxillary, mediastinal, and retroperitoneal lymph nodes; oral cavity; gastrointestinal system; skin; and breasts. Sites of metastasis include lung and liver. Immunophenotype includes CD21, CD35, and CD23. Clinical course is typically indolent, and treatment includes surgical excision followed by regional radiotherapy and sometimes systemic chemotherapy.

Hemophagocytic Syndrome Hemophagocytic syndrome (HPS), also known as hemophagocytic lymphohistiocytosis (HLH), represents nonneoplastic proliferation and activation of macrophages that induce cytokine-mediated BM suppression and features of intense phagocytosis in BM and liver. HPS may result from genetic (primary) or acquired (secondary) disorders of macrophages. The former entail genetically determined inability to regulate macrophage proliferation and activation and might include alterations in familial HLH genes (*PRF1*, *UNC13D*, *STXBP2*, and *STX11*), granule/pigment abnormality genes (*RAB27A*, *LYST*, and *AP3B1*), or X-linked lymphoproliferative disease genes (*SH2D1A* and *XIAP*). Acquired HPS is often precipitated by viral infections, most notably Epstein-Barr virus. HPS might also accompany certain malignancies such as T-cell lymphoma and autoimmune diseases. Clinical presentation includes fever, severe constitutional symptoms, enlarged lymph nodes, hepatosplenomegaly, neurologic dysfunction, and abnormalities in multiple organ function tests. Diagnosis is accomplished by either detection of HLH-related mutations or meeting five of the following eight conventional criteria: (1) hemophagocytosis in the BM/spleen/lymph nodes; (2) serum ferritin $\geq 500\text{ }\mu\text{g/L}$; (3) hypofibrinogenemia (fibrinogen $\leq 1.5\text{ g/L}$) or hypertriglyceridemia (triglycerides $\geq 3\text{ mmol/L}$); (4) low NK cell activity; (5) elevated soluble IL-2 receptor (CD25) $\geq 2400\text{ U/mL}$; (6) bi- or tricytopenia (platelets $<100 \times 10^9/\text{L}$, hemoglobin $<9\text{ g/dL}$, absolute neutrophil count $<1 \times 10^9/\text{L}$); (7) splenomegaly palpable $>3\text{ cm}$ below left costal margin; and (8) fever. Clinical course is often fulminant and fatal. Current therapeutic approaches for primary or secondary HLH include the so-called "HLH-94 protocol," which consists of weekly treatments with etoposide and dexamethasone, stem cell transplantation, emapalumab (a monoclonal antibody that binds and neutralizes interferon γ), and the JAK1/2 inhibitor ruxolitinib. Emapalumab was FDA approved in

November 2018 for use in pediatric and adult patients with primary HLH with refractory or progressive disease.

FURTHER READING

- ALLEN CE et al: The coming of age of Langerhans cell histiocytosis. *Nat Immunol* 21:1, 2020.
- DAO KT et al: Efficacy of ruxolitinib in patients with chronic neutrophilic leukemia and atypical chronic myeloid leukemia. *J Clin Oncol* 38:1006, 2020.
- MAURER M et al: Results from PIONEER: A randomized, double-blind, placebo-controlled, phase 2 study of avapritinib in patients with indolent systemic mastocytosis (ISM). *Oncol Res Treat* 43:77, 2020.
- PATNAIK MM, TEFFERI A: Chronic myelomonocytic leukemia: 2020 update on diagnosis, risk stratification and management. *Am J Hematol* 95:97, 2020.
- ROUFSSE F et al: Efficacy and safety of mepolizumab in hypereosinophilic syndrome: A phase III, randomized, placebo-controlled trial. *J Allergy Clin Immunol* 146:1397, 2020.
- SUKSWAI N et al: Diffuse large B-cell lymphoma variants: An update. *Pathology* 52:53, 2020.
- SWERDLOW SH et al: The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 127:2375, 2016.
- SZUBER N et al: Chronic neutrophilic leukemia: 2020 update on diagnosis, molecular genetics, prognosis, and management. *Am J Hematol* 95:212, 2020.
- VALLURUPALLI M, BERLINER N: Emalumab for the treatment of relapsed/refractory hemophagocytic lymphohistiocytosis. *Blood* 134:1783, 2019.

111

Plasma Cell Disorders

Nikhil C. Munshi, Dan L. Longo,
Kenneth C. Anderson



The *plasma cell disorders* are monoclonal neoplasms related to each other by virtue of their development from common progenitors in the late B-lymphocyte lineage. Multiple myeloma (MM), Waldenström's macroglobulinemia, primary amyloidosis ([Chap. 112](#)), and the heavy chain diseases comprise this group and may be designated by a variety of synonyms such as *monoclonal gammopathies*, *paraproteinemias*, *plasma cell dyscrasias*, and *dysproteinemias*. Mature B lymphocytes destined to produce IgG bear surface immunoglobulin molecules of both μ and γ heavy chain isotypes with both isotypes having identical idiotypes (variable regions). Under normal circumstances, maturation to antibody-secreting plasma cells and their proliferation is stimulated by exposure to the antigen for which the surface immunoglobulin is specific; however, in the plasma cell disorders, the control over this process is lost. The clinical manifestations of all the plasma cell disorders relate to the expansion of the neoplastic cells, to the secretion of cell products (immunoglobulin molecules or subunits, lymphokines), and to some extent to the host's response to the tumor. Normal development of B lymphocytes is discussed in [Chap. 349](#) and depicted in [Fig. 108-2](#).

Three categories of structural variation are present among immunoglobulin molecules that form antigenic determinants, and these are used to classify immunoglobulins. *Isotypes* are those determinants that distinguish among the main classes of antibodies of a given species and are the same in all normal individuals of that species. Therefore, isotypic determinants are, by definition, recognized by antibodies from a distinct species (heterologous sera) but not by antibodies from the same species (homologous sera). There are five heavy chain isotypes (M, G, A, D, E) and two light chain isotypes (κ , λ). *Allotypes* are distinct determinants that reflect regular small differences between individuals

of the same species in the amino acid sequences of otherwise similar immunoglobulins. These differences are determined by allelic genes; by definition, they are detected by antibodies made in the same species. *Idiotypes* are the third category of antigenic determinants. They are unique to the molecules produced by a given clone of antibody-producing cells. Idiotypes are formed by the unique structure of the antigen-binding portion of the molecule.

Antibody molecules ([Fig. 111-1](#)) are composed of two heavy chains (~50,000 molecular weight [mol wt]) and two light chains (~25,000 mol wt). Each chain has a constant portion (limited amino acid sequence variability) and a variable region (extensive sequence variability). The light and heavy chains are linked by disulfide bonds and are aligned so that their variable regions are adjacent to one another. This variable region forms the antigen recognition site of the antibody molecule; its unique structural features form idiotypes that are reliable markers for a particular clone of cells because each antibody is formed and secreted by a single clone. Because of the mechanics of the gene rearrangements necessary to specify the immunoglobulin variable regions (VDJ joining for the heavy chain, VJ joining for the light chain), a particular clone rearranges only one of the two chromosomes to produce an immunoglobulin molecule of only one light chain isotype and only one allotype (allelic exclusion) ([Fig. 111-1](#)). After exposure to antigen, the variable region may become associated with a new heavy chain isotype (class switch). Each clone of cells performs these sequential gene arrangements in a unique way. This results in each clone producing a unique immunoglobulin molecule. In most plasma cells, light chains are synthesized in slight excess, secreted as free light chains, and cleared by the kidney, but <10 mg of such light chains is excreted per day.

Electrophoretic analysis permits separation of components of the serum proteins ([Fig. 111-2](#)). The immunoglobulins move heterogeneously in an electric field and form a broad peak in the gamma region, which is usually increased in the sera of patients with plasma cell tumors. There is a sharp spike in this region called an *M component* (M for monoclonal). Less commonly, the M component may appear in the β_2 or α_2 globulin region. The monoclonal antibody must be present at a concentration of at least 5 g/L (0.5 g/dL) to be accurately quantitated by this method. This corresponds to $\sim 10^9$ cells producing the antibody. Confirmation of the type of immunoglobulin and that it is truly monoclonal is determined by immunoelectrophoresis that reveals a single heavy and/or light chain type. Hence, immunoelectrophoresis and electrophoresis provide qualitative and quantitative assessment of the M component, respectively. Once the presence of an M component has been confirmed, the amount of M component in the serum is a reliable measure of the tumor burden, making M component an excellent tumor marker to manage therapy, yet it is not specific enough to be used to screen asymptomatic patients. In addition to the plasma cell disorders, M components may be detected in other lymphoid neoplasms such as chronic lymphocytic leukemia (CLL) and lymphomas of B- or T-cell origin; nonlymphoid neoplasms such as chronic myeloid leukemia, breast cancer, and colon cancer; a variety of nonneoplastic conditions such as cirrhosis, sarcoidosis, parasitic diseases, Gaucher's disease, and pyoderma gangrenosum; and a number of autoimmune conditions, including rheumatoid arthritis, myasthenia gravis, and cold agglutinin disease. Monoclonal proteins are also observed in immunosuppressed patients after organ transplant and, rarely, allogeneic transplant. At least two very rare skin diseases—lichen myxedematosus (also known as papular mucinosis) and necrobiotic xanthogranuloma—are associated with a monoclonal gammopathy. In papular mucinosis, highly cationic IgG is deposited in the dermis of patients. This organ specificity may reflect the specificity of the antibody for some antigenic component of the dermis. Necrobiotic xanthogranuloma is a histiocytic infiltration of the skin, usually of the face, that produces red or yellow nodules that can enlarge to plaques. Approximately 10% progress to myeloma. Five percent of patients with sensory motor neuropathy also have a monoclonal paraprotein.

The nature of the M component is variable in plasma cell disorders. It may be an intact antibody molecule of any heavy chain subclass, or it may be an altered antibody or fragment. Isolated light or heavy chains may be produced. In some plasma cell tumors such as extramedullary

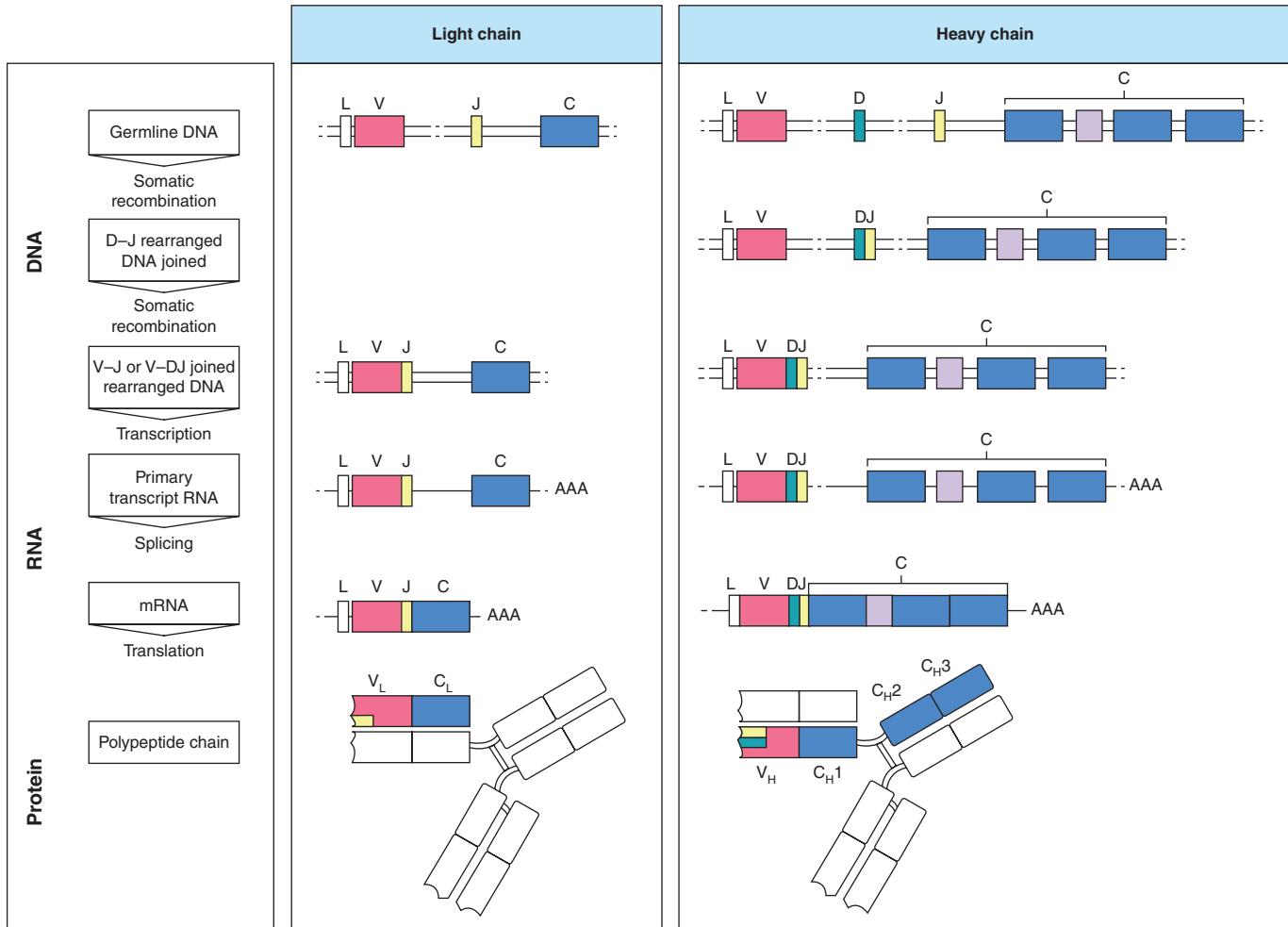
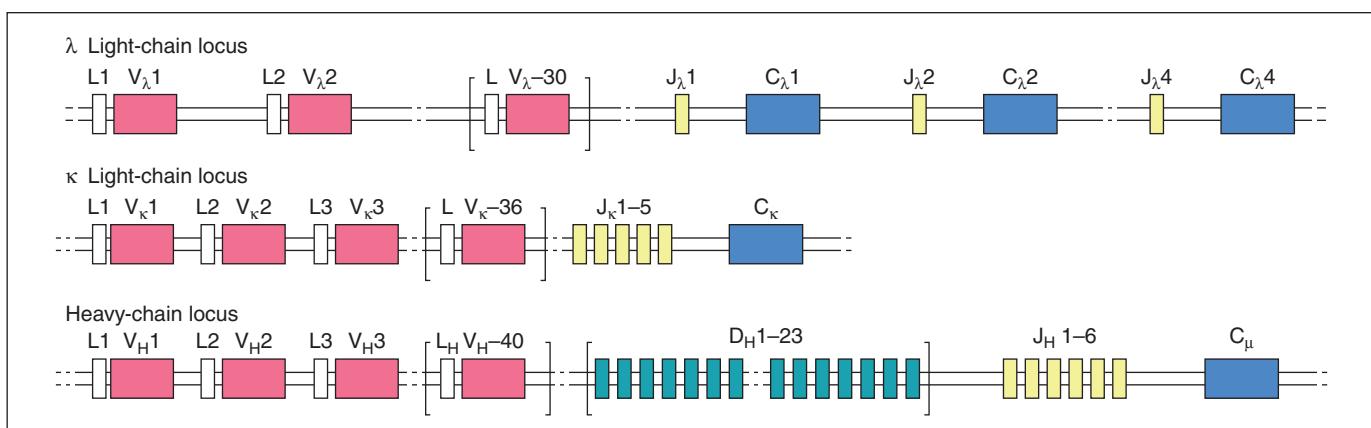


FIGURE 111-1 Immunoglobulin genetics and the relationship of gene segments to the antibody protein. The top portion of the figure is a schematic of the organization of the immunoglobulin genes, λ on chromosome 22, κ on chromosome 2, and the heavy chain locus on chromosome 14. The heavy chain locus is >2 megabases, and some of the D region gene segments are only a few bases long, so the figure depicts the schematic relationship among the segments, not their actual size. The bottom portion of the figure outlines the steps in going from the noncontiguous germline gene segments to an intact antibody molecule. Two recombination events juxtapose the V-D-J (or V-J for light chains) segments. The rearranged gene is transcribed, and RNA splicing cuts out intervening sequences to produce an mRNA, which is then translated into an antibody light or heavy chain. The sites on the antibody that bind to antigen (the so-called CDR3 regions) are encoded by D and J segments for heavy chains and the J segments for light chains. (From Janeway's Immunobiology, 9th ed by Kenneth Murphy and Casey Weaver. Copyright © 2017 by Garland Science, Taylor & Francis Group, LLC. Used by permission of W. W. Norton & Company, Inc.)

or solitary bone plasmacytomas, less than one-third of patients will have an M component. In ~20% of myelomas, only light chains are produced and, in most cases, are secreted in the urine as Bence Jones proteins. The frequency of myelomas of a particular heavy chain class is roughly proportional to the serum concentration, and therefore, IgG myelomas are more common than IgA and IgD myelomas. In ~1% of patients with myeloma, byclonal or triclonal gammopathy is observed.

MULTIPLE MYELOMA

■ DEFINITION

MM represents a malignant proliferation of plasma cells derived from a single clone. The tumor, its products, and the host response to it result in a number of organ dysfunctions and symptoms, including bone pain or fracture, renal failure, susceptibility to infection, anemia,

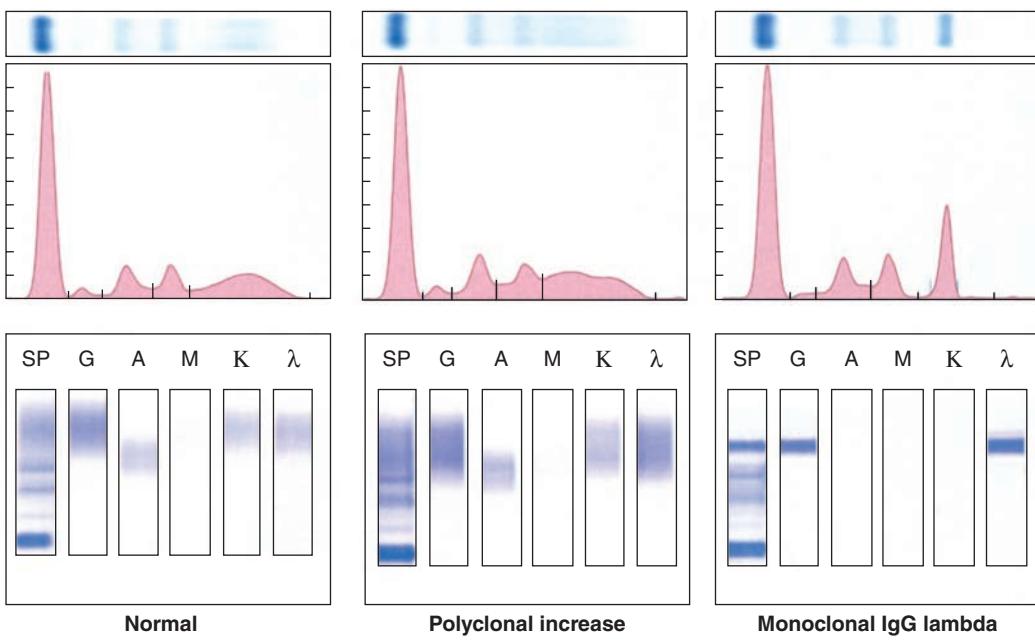


FIGURE 111-2 Representative patterns of serum electrophoresis and immunofixation. The upper panels represent agarose gel, middle panels are the densitometric tracing of the gel, and lower panels are immunofixation patterns. The panel on the left illustrates the normal pattern of serum protein on electrophoresis. Because there are many different immunoglobulins in the serum, their differing mobilities in an electric field produce a broad peak. In conditions associated with increases in polyclonal immunoglobulin, the broad peak is more prominent (middle panel). In monoclonal gammopathies, the predominance of a product of a single cell produces a “church spire” sharp peak, usually in the γ globulin region (right panel). The immunofixation (lower panel) identifies the type of immunoglobulin. For example, normal and polyclonal increases in immunoglobulins produce no distinct bands; however, the right panel shows distinct bands in IgG and lambda protein lanes, confirming the presence of IgG lambda monoclonal protein. (Courtesy of Dr. Neal I. Lindeman.)

hypercalcemia, and occasionally clotting abnormalities, neurologic symptoms, and manifestations of hyperviscosity.

ETIOLOGY

The cause of myeloma is not known. Myeloma occurred with increased frequency in those exposed to the radiation of nuclear warheads in World War II after a 20-year latency. Myeloma has been seen more commonly than expected among farmers, wood workers, leather workers, and those exposed to petroleum products. A variety of recurrent chromosomal alterations have been found in patients with myeloma: hyperdiploidy (trisomies involving one or more of chromosomes 3, 5, 7, 9, 11, 15, 19, or 21) is observed in half of the patients, while the other half have translocations involving the 14q32 chromosome with variable partners including t(11;14)(q13;q32), t(4;14)(p16;q32), and t(14;16). Other frequent abnormalities include 13q14 deletion, 1q amplification or 1p deletion, and 17p13 deletions. Evidence is strong that errors in switch recombination—the genetic mechanism to change antibody heavy chain isotype—participate in the early transformation process. However, no single common molecular pathogenetic pathway has yet emerged. Genome sequencing studies have failed to identify any recurrent mutation with frequency >20%; N-ras, K-ras, and B-raf mutations are most common and, combined, occur in >40% of patients. Evidence of complex clusters of subclonal variants is present at diagnosis, and additional mutations are acquired over time, indicative of genomic evolution that may drive disease progression. Interleukin (IL) 6 may play a role in driving myeloma cell proliferation. It remains difficult to distinguish benign from malignant plasma cells based on morphologic criteria in all but a few cases (Fig. 111-3).

INCIDENCE AND PREVALENCE

In 2021 in the United States, 34,920 new cases of myeloma were estimated to be diagnosed, and 12,410 people were estimated to die from the disease. Myeloma increases in incidence with age. The median age at diagnosis is 69 years; it is uncommon under age 40. Males are more commonly affected than females, and blacks have nearly twice the incidence of whites. In 2018, myeloma accounted for 1.8% of all malignancies, with incidence rates per 100,000 of 6.1 in whites and 13.6 in blacks.

GLOBAL CONSIDERATIONS

The incidence of myeloma is highest in blacks and Pacific Islanders; intermediate in Europeans and North American whites; and lowest in people from developing countries including Asia. The higher incidence in more developed countries may result from the combination of a longer life expectancy and more frequent medical surveillance. Incidence of MM in other ethnic groups including native Hawaiians, female Hispanics, American Indians from New Mexico, and Alaskan natives is higher relative to U.S. whites in the same geographic area. Chinese and Japanese populations have a lower incidence than whites. Immunoproliferative small-intestinal disease (IPSID) with a heavy chain disease is most prevalent in the Mediterranean area. Despite these differences in prevalence, the characteristics, response to therapy, and prognosis of myeloma are similar worldwide.

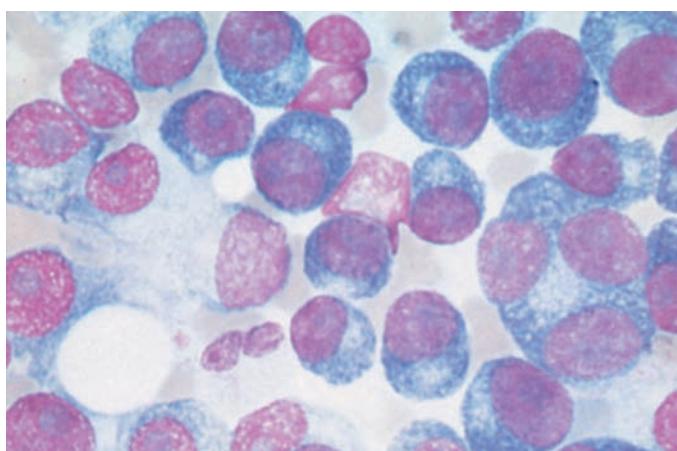


FIGURE 111-3 Multiple myeloma (marrow). The cells bear characteristic morphologic features of plasma cells: round or oval cells with an eccentric nucleus composed of coarsely clumped chromatin, a densely basophilic cytoplasm, and a perinuclear clear zone containing the Golgi apparatus. Binucleate and multinucleate malignant plasma cells can be seen.

PATHOGENESIS AND CLINICAL MANIFESTATIONS

MM cells bind via cell-surface adhesion molecules to bone marrow stromal cells (BMSCs) and extracellular matrix (ECM), which triggers MM cell growth, survival, drug resistance, and migration in the bone marrow milieu (Fig. 111-4). These effects are due both to direct MM cell–BMSC binding via adhesion molecules and to induction of various cytokines, including IL-6, insulin-like growth factor type 1 (IGF-1), vascular endothelial growth factor (VEGF), and stromal cell-derived growth factor (SDF)-1 α . Growth, drug resistance, and migration are mediated via Ras/Raf/mitogen-activated protein kinase, PI3K/Akt, and protein kinase C signaling cascades, respectively. Other cellular elements in the bone marrow microenvironment also significantly impact MM cell growth and survival. The major myeloma supporting interactions are with endothelial cells and osteoclasts. Immune cells such as plasmacytoid dendritic cells (pDC), myeloid-derived suppressor cells (MDSC), and T helper 17 (T_H17) cells are increased in number and support myeloma growth, while antimyeloma immune responses, especially T helper and cytotoxic cells, B cells, and natural killer T cells, are suppressed.

Bone pain is the most common symptom in myeloma, affecting nearly 70% of patients. Persistent localized pain usually signifies a pathologic fracture. The bone lesions of myeloma are caused by the proliferation of tumor cells, activation of osteoclasts that destroy bone, and suppression of osteoblasts that form new bone. The increased osteoclast activity is mediated by osteoclast activating factors (OAFs) produced by the myeloma cells (mediated by several cytokines, including IL-1, lymphotoxin, vascular endothelial growth factor [VEGF], receptor activator of nuclear factor- κ B [RANK] ligand, macrophage

inhibitory factor [MIP]-1 α , and tumor necrosis factor [TNF]). The bone lesions are lytic in nature (Fig. 111-5) and are rarely associated with osteoblastic new bone formation due to their suppression by dickhoff-1 (DKK-1) produced by myeloma cells. Therefore, radioisotopic bone scanning is less useful in diagnosis than is plain radiography. The bony lysis results in substantial mobilization of calcium from bone, and serious acute and chronic complications of hypercalcemia may dominate the clinical picture (see below). Localized bone lesions may cause the collapse of vertebrae, leading to spinal cord compression. The next most common clinical problem in patients with myeloma is susceptibility to bacterial infections. The most common infections are pneumonias and pyelonephritis, and the most frequent pathogens are *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Klebsiella pneumoniae* in the lungs and *Escherichia coli* and other gram-negative organisms in the urinary tract. In ~25% of patients, recurrent infections are the presenting features, and >75% of patients will have a serious infection at some time in their course. The susceptibility to infection has several contributing causes. First, patients with myeloma have diffuse hypogammaglobulinemia if the M component is excluded. The hypogammaglobulinemia is related to both decreased production and increased destruction of normal antibodies. The large M component results in fractional catabolic rates of 8–16% instead of the normal 2%. Moreover, some patients generate a population of circulating regulatory cells in response to their myeloma that can suppress normal antibody synthesis. These patients have very poor antibody responses, especially to polysaccharide antigens such as those on bacterial cell walls. Various abnormalities in T-cell function are also observed including decreased T_H1 response, increase in T_H17 cells producing

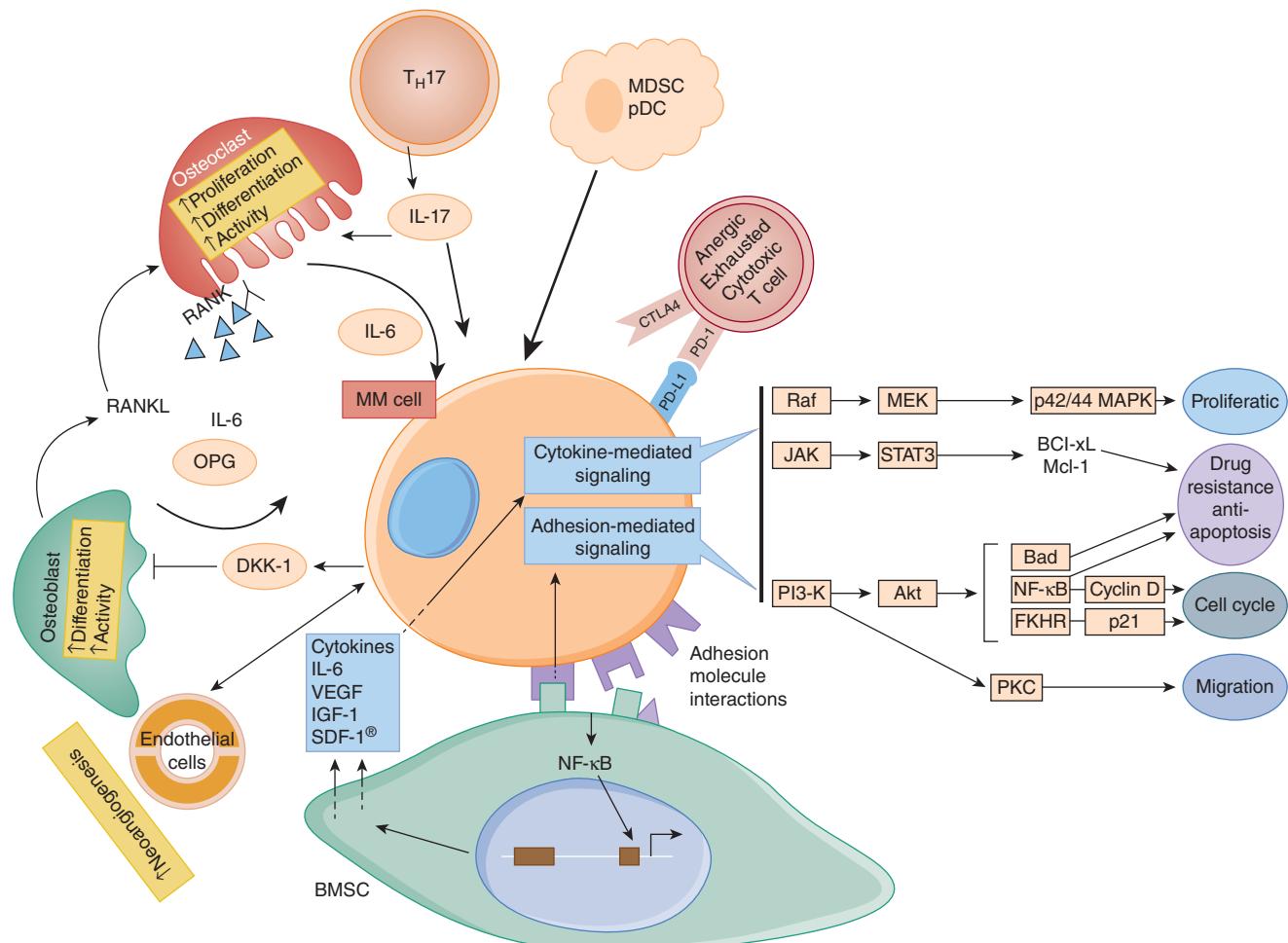


FIGURE 111-4 Pathogenesis of multiple myeloma. Multiple myeloma (MM) cells interact with bone marrow stromal cells (BMSCs) and extracellular matrix proteins via adhesion molecules, triggering adhesion-mediated signaling as well as cytokine production. This triggers cytokine-mediated signaling that provides growth, survival, and antiapoptotic effects as well as development of drug resistance. Additional bidirectional interactions lead to inhibition of osteoblast and increase in osteoclast activity which leads to bone-related issues in myeloma. Similar interactions with immune microenvironment lead to augmentation of tumor promoting immune responses and suppression of tumor protective immune responses, overall allowing myeloma cell growth. (Adapted from G Bianchi, NC Munshi: Blood 125: 3049, 2015.)



A



B

FIGURE 111-5 Bony lesions in multiple myeloma (MM). **A.** The skull demonstrates the typical “punched out” lesions characteristic of MM. The lesion represents a purely osteolytic lesion with little or no osteoblastic activity (above). **B.** PET/CT showing multiple fluorodeoxyglucose (FDG)-avid lesions in skeleton (left panel) with their resolution on achieving complete response (CR) (right panel). (Part A courtesy of Dr. Geraldine Schechter; with permission. Part B courtesy of Dr. Sundar Jagannath; with permission.)

proinflammatory cytokines, and aberrant T regulatory cell function. Granulocyte lysozyme content is low, and granulocyte migration is not as rapid as normal in patients with myeloma, probably the result of a tumor product. There are also a variety of abnormalities in complement functions in myeloma patients. All these factors contribute to the immune deficiency in these patients. Some commonly used therapeutic agents may significantly affect immune function; e.g., dexamethasone suppresses immune responses and increases susceptibility to bacterial and fungal infection, B-cell maturation antigen (BCMA)-targeting chimeric antigen receptor T (CAR-T) cells can eliminate plasma cells inducing hypogammaglobulinemia, and bortezomib predisposes to herpesvirus reactivation.

Renal failure occurs in nearly 25% of myeloma patients, and some renal pathology is noted in >50%. Of many contributing factors, hypercalcemia is the most common cause of renal failure. Glomerular deposits of amyloid, hyperuricemia, recurrent infections, frequent use of nonsteroidal anti-inflammatory agents for pain control, use of iodinated contrast dye for imaging, bisphosphonate use, and occasional infiltration of the kidney by myeloma cells all may contribute to renal dysfunction. However, tubular damage associated with the excretion of light chains is almost always present. Normally, light chains are filtered, reabsorbed in the tubules, and catabolized. With the increase in the amount of light chains presented to the tubule, the tubular cells become overloaded with these proteins, and tubular damage results either directly from light chain toxic effects or indirectly from the release of intracellular lysosomal enzymes. The earliest manifestation of this tubular damage is the adult Fanconi’s syndrome (a type 2 proximal renal tubular acidosis), with loss of glucose and amino acids, as well as defects in the ability of the kidney to acidify and concentrate the urine. The proteinuria is not accompanied by hypertension, and the protein is nearly all light chains. Generally, very little albumin is in the urine because glomerular function is usually normal. When the glomeruli are involved, nonselective proteinuria is also observed. Patients with myeloma also have a decreased anion gap [i.e., $\text{Na}^+ - (\text{Cl}^- + \text{HCO}_3^-)$] because the M component is cationic, resulting in retention of chloride. This is often accompanied by hyponatremia that is felt to be artificial (pseudohyponatremia) because each volume of serum has less water as a result of the increased protein. Renal dysfunction due to light chain deposition disease, light chain cast nephropathy, and amyloidosis is partially reversible with effective therapy. Myeloma patients are susceptible to developing acute renal failure if they become dehydrated.

Normocytic and normochromic anemia occurs in ~80% of myeloma patients. It is usually related to the replacement of normal marrow by expanding tumor cells, to the inhibition of hematopoiesis by factors made by the tumor, to reduced production of erythropoietin by the kidney, and to the effects of long-term therapy. In addition, mild hemolysis may contribute to the anemia. A larger than expected fraction of patients may have megaloblastic anemia due to either folate or vitamin B₁₂ deficiency. Granulocytopenia and thrombocytopenia are rare except when therapy-induced. Clotting abnormalities may be seen due to the failure of antibody-coated platelets to function properly; the interaction of the M component with clotting factors I, II, V, VII, or VIII; antibody to clotting factors; or amyloid damage of endothelium. Deep venous thrombosis is also observed with use of thalidomide, lenalidomide, or pomalidomide in combination with dexamethasone. Raynaud’s phenomenon and impaired circulation may result if the M component forms cryoglobulins, and hyperviscosity syndromes may develop depending on the physical properties of the M component (most common with IgM, IgG3, and IgA paraproteins). Hyperviscosity is defined based on the relative viscosity of serum as compared with water. Normal relative serum viscosity is 1.8 (i.e., serum is normally almost twice as viscous as water). Symptoms of hyperviscosity occur at a level greater than 4 centipoises (cP), which is usually reached at paraprotein concentrations of ~40 g/L (4 g/dL) for IgM, 50 g/L (5 g/dL) for IgG3, and 70 g/L (7 g/dL) for IgA; however, depending on chemical and physical properties of the paraprotein molecule, it can occasionally be observed at lower levels.

Although neurologic symptoms occur in a minority of patients, they may have many causes. Hypercalcemia may produce lethargy, weakness, depression, and confusion. Hyperviscosity may lead to headache, fatigue, shortness of breath, exacerbation or precipitation of heart failure, visual disturbances, ataxia, vertigo, retinopathy, somnolence, and coma. Bony damage and collapse may lead to cord compression, radicular pain, and loss of bowel and bladder control. Infiltration of peripheral nerves by amyloid can be a cause of carpal tunnel syndrome and other sensorimotor mono- and polyneuropathies. Neuropathy associated with monoclonal gammopathy of undetermined significance (MGUS) and myeloma is more frequently sensory than motor neuropathy and is associated with IgM more than other isotypes. In >50% of patients with neuropathy, the IgM monoclonal protein is directed against myelin-associated globulin (MAG). Sensory neuropathy is also a side effect of therapy, specifically thalidomide and bortezomib.

Many of the clinical features of myeloma, e.g., cord compression, pathologic fractures, hyperviscosity, sepsis, and hypercalcemia, can present as medical emergencies. Despite the widespread distribution of plasma cells in the body, tumor expansion is dominantly within bone and bone marrow and, for reasons unknown, rarely causes enlargement of spleen, lymph nodes, or gut-associated lymphatic tissue.

■ DIAGNOSIS AND STAGING

The diagnosis of myeloma requires marrow plasmacytosis ($>10\%$), a serum and/or urine M component, and at least one of the myeloma-defining events detailed in **Table 111-1**. Bone marrow plasma cells are CD138+ and either monoclonal kappa or lambda light chain positive. The most important differential diagnosis in patients with myeloma involves their separation from individuals with MGUS or smoldering multiple myeloma (SMM). MGUS is vastly more common than myeloma, occurring in 1% of the population aged >50 years and in up to 10% of individuals aged >75 years. The diagnostic criteria for

MGUS, SMM, and myeloma are described in Table 111-1. Although $\sim 1\%$ of patients per year with MGUS go on to develop myeloma, all cases of myeloma are preceded by MGUS. Non-IgG subtype, abnormal kappa/lambda free light chain ratio, and serum M protein >15 g/L (1.5 g/dL) are associated with higher incidence of progression of MGUS to myeloma. Absence of all three features predicts a 5% chance of progression, whereas higher-risk MGUS with the presence of all three features predicts a 60% chance of progression over 20 years. The features responsible for higher risk of progression from SMM to MM are bone marrow plasmacytosis $>10\%$, abnormal kappa/lambda free light chain ratio, and serum M protein >30 g/L (3 g/dL). Patients with only one of these three features have a 25% chance of progression to MM in 5 years, whereas patients with high-risk SMM with all three features have a 76% chance of progression. Two important variants of myeloma are solitary bone plasmacytoma and solitary extramedullary plasmacytoma. These lesions are associated with an M component in $<30\%$ of the cases, they may affect younger individuals, and both are associated

TABLE 111-1 Diagnostic Criteria for Multiple Myeloma, Myeloma Variants, and Monoclonal Gammopathy of Undetermined Significance

Monoclonal Gammopathy of Undetermined Significance (MGUS)

Serum monoclonal protein (non-IgM type) <30 g/L

Clonal bone marrow plasma cells $<10\%^a$

Absence of myeloma-defining events or amyloidosis that can be attributed to the plasma cell proliferative disorder

Smoldering Multiple Myeloma (Asymptomatic Myeloma)

Both criteria must be met:

- Serum monoclonal protein (IgG or IgA) ≥ 30 g/L or urinary monoclonal protein ≥ 500 mg per 24 h and/or clonal bone marrow plasma cells 10–60%
- Absence of myeloma-defining events or amyloidosis

Symptomatic Multiple Myeloma

Clonal bone marrow plasma cells or biopsy-proven bony or extramedullary plasmacytoma^a and any one or more of the following myeloma-defining events:

- Evidence of one or more indicators of end-organ damage that can be attributed to the underlying plasma cell proliferative disorder, specifically:
 - Hypercalcemia: serum calcium >0.25 mmol/L (>1 mg/dL) higher than the upper limit of normal or >2.75 mmol/L (>11 mg/dL)
 - Renal insufficiency: creatinine clearance <40 mL/min^b or serum creatinine >177 μ mol/L (>2 mg/dL)
 - Anemia: hemoglobin value <20 g/L below the lower limit of normal, or a hemoglobin value <100 g/L
 - Bone lesions: one or more osteolytic lesions on skeletal radiography, CT, or PET-CT^c
 - Any one or more of the following biomarkers of malignancy:
 - Clonal bone marrow plasma cell percentage^a $\geq 60\%$
 - Involved: uninvolved serum free light chain ratio^d ≥ 100
 - >1 focal lesion on MRI studies^e

Nonsecretory Myeloma

No M protein in serum and/or urine with immunofixation^f

Bone marrow clonal plasmacytosis $\geq 10\%$ or plasmacytoma^a

Myeloma-related organ or tissue impairment (end-organ damage, as described above)

Solitary Plasmacytoma

Biopsy-proven solitary lesion of bone or soft tissue with evidence of clonal plasma cells

Normal bone marrow with no evidence of clonal plasma cells^a

Normal skeletal survey and MRI (or CT) of spine and pelvis (except for the primary solitary lesion)

Absence of end-organ damage such as hypercalcemia, renal insufficiency, anemia, or bone lesions (CRAB) that can be attributed to a lymphoplasma cell proliferative disorder

POEMS Syndrome

All of the following four criteria must be met:

1. Polyneuropathy
2. Monoclonal plasma cell proliferative disorder
3. Any one of the following: (a) sclerotic bone lesions; (b) Castleman's disease; (c) elevated levels of vascular endothelial growth factor (VEGF)
4. Any one of the following: (a) organomegaly (splenomegaly, hepatomegaly, or lymphadenopathy); (b) extravascular volume overload (edema, pleural effusion, or ascites); (c) endocrinopathy (adrenal, thyroid, pituitary, gonadal, parathyroid, and pancreatic); (d) skin changes (hyperpigmentation, hypertrichosis, glomeruloid hemangioma, plethora, acrocyanosis, flushing, and white nails); (e) papilledema; (f) thrombocytosis/polycythemia^g

^aClonality should be established by showing κ/λ light chain restriction on flow cytometry, immunohistochemistry, or immunofluorescence. Bone marrow plasma cell percentage should preferably be estimated from a core biopsy specimen; in case of a disparity between the aspirate and core biopsy, the highest value should be used. ^bMeasured or estimated by validated equations. ^cIf bone marrow has $<10\%$ clonal plasma cells, more than one bone lesion is required to distinguish from solitary plasmacytoma with minimal marrow involvement. ^dThese values are based on the serum FreeLight assay (The Binding Site Group, Birmingham, United Kingdom). The involved free light chain must be ≥ 100 mg/L. ^eEach focal lesion must be ≥ 5 mm in size. ^fA small M component may sometimes be present. ^gThese features should have no other attributable causes and have temporal relation with each other.

Abbreviations: PET-CT, ¹⁸F-fluorodeoxyglucose positron emission tomography with computed tomography; POEMS, polyneuropathy, organomegaly, endocrinopathy, M-protein, and skin changes.

with median survivals of ≥ 10 years. Solitary bone plasmacytoma is a single lytic bone lesion without marrow plasmacytosis. Extramedullary plasmacytomas usually involve the submucosal lymphoid tissue of the nasopharynx or paranasal sinuses without marrow plasmacytosis. Both tumors are highly responsive to local radiation therapy. If an M component is present, it should disappear after treatment. Solitary bone plasmacytomas may recur in other bony sites or evolve into myeloma. Extramedullary plasmacytomas rarely recur or progress.

Serum protein electrophoresis and measurement of serum immunoglobulins and free light chains are useful for detecting and characterizing M spikes, supplemented by immunoelectrophoresis, which is especially sensitive for identifying low concentrations of M components not detectable by protein electrophoresis. A 24-h urine specimen is necessary to quantitate Bence Jones protein (immunoglobulin light chain) excretion. The serum M component will be IgG in 53% of patients, IgA in 25%, and IgD in 1%; 20% of patients will have only light chains in serum and urine. Dipsticks for detecting proteinuria are not reliable at identifying light chains, and the heat test for detecting Bence Jones protein is falsely negative in $\sim 50\%$ of patients with light chain myeloma. Fewer than 1% of patients have no identifiable M component; these patients usually have light chain myeloma in which renal catabolism has made the light chains undetectable in the urine. In most of these patients, light chains can now be detected by serum free light chain assay. IgD myeloma may also present with light chain disease. About two-thirds of patients with serum M components also have urinary light chains. The light chain isotype may have an impact on disease behavior. Whether this is due to some genetically important determinant of cell proliferation or because lambda light chains are more likely to cause renal damage and form amyloid than are kappa light chains is unclear. The heavy chain isotype may have an impact on patient management as well. About half of patients with IgM paraproteins develop hyperviscosity compared with only 2–4% of patients with IgA and IgG M components. Among IgG myelomas, it is the IgG3 subclass that has the highest tendency to form both concentration- and temperature-dependent aggregates, leading to hyperviscosity and cold agglutination at lower serum concentrations. A standard workup directed at detecting monoclonal plasma cells and myeloma-defining events as well as prognosis is detailed in **Table 111-2**.

A complete blood count with differential may reveal anemia. Erythrocyte sedimentation rate is elevated. Rare patients ($\sim 1\%$) may have plasma cell leukemia with >2000 plasma cells/ μL . This may be seen in disproportionate frequency in IgD (12%) and IgE (25%) myelomas. Serum calcium, urea nitrogen, creatinine, and uric acid levels may be elevated. Serum alkaline phosphatase is usually normal even with extensive bone involvement because of the absence of osteoblastic activity. It is also important to quantitate serum β_2 -microglobulin and albumin (see below).

Chest and bone radiographs may reveal lytic lesions or diffuse osteopenia. Magnetic resonance imaging (MRI) offers a sensitive means to document extent of bone marrow infiltration and cord or root compression in patients with pain syndromes. ^{18}F -fluorodeoxyglucose (^{18}F -FDG) positron emission tomography (PET)/computed tomography (CT) is a valuable tool to assess bone damage and detect extramedullary sites of the disease (Fig. 111-5). The use of ^{18}F -FDG PET/CT is recommended to distinguish between smoldering and active MM and to confirm a suspected diagnosis of solitary plasmacytoma. It is also a valuable tool to evaluate response in patients with oligo- or nonsecretory myeloma.

PROGNOSIS

Serum β_2 -microglobulin is the single most powerful predictor of survival and can substitute for staging. β_2 -Microglobulin is the light chain of the class I major histocompatibility antigens (HLA-A, -B, -C) on the surface of every cell. Combination of serum β_2 -microglobulin and albumin levels forms the basis for a three-stage International Staging System (ISS) (**Table 111-3**) that predicts survival. With the use of high-dose therapy and the newer agents, the Durie-Salmon staging system is unable to predict outcome and is no longer used. High labeling index, circulating plasma cells, performance status, and high levels of lactate dehydrogenase are also associated with poor prognosis.

TABLE 111-2 Standard Investigative Workup in Multiple Myeloma (MM)

Investigations to Evaluate for Clonal Plasma Cells

- Bone marrow aspirate and biopsy (fine-needle aspiration of plasmacytoma if indicated)
- Histology
- Clonality by kappa/lambda immunostaining by flow cytometry or immunohistochemistry

Investigations to Evaluate Clonal Paraprotein

- Serum protein electrophoresis and immunofixation
- Quantitative serum immunoglobulin levels (IgG, IgA, and IgM)
- 24-h urine protein electrophoresis and immunofixation
- Serum free light chain and ratio
- Immunofixation for IgD or IgE in select cases

Investigation to Evaluate End-Organ Damage

- Hemogram to assess for anemia
- Chemistry panel for renal function and calcium
- Skeletal survey to evaluate bone lesions
- PET/CT or MRI if smoldering MM or solitary plasmacytoma with no other MDE or extramedullary disease

Investigation for Risk Stratification

- β_2 -Microglobulin and serum albumin for ISS stage
- Fluorescent in situ hybridization for hyperdiploidy, del17p, t(4;14); t(11;14), t(14;16), t(14;20), amp1q34, and del13 on bone marrow sample
- LDH

Specialized Investigation in Selected Cases

- Abdominal fat pad for amyloid
- Serum viscosity if IgM component or high IgA levels or serum M component >7 g/dL
- Myd88* and *CXCR4* mutation analysis if IgM component

Abbreviations: ISS, International Staging System; LDH, lactate dehydrogenase; MRI, magnetic resonance imaging; PET/CT, positron emission tomography/computed tomography.

Other factors that may influence prognosis are detection of any cytogenetic abnormalities including hypodiploidy by karyotype, fluorescent in situ hybridization (FISH)-identified chromosome 17p deletion, and translocations t(4;14), (14;16), and t(14;20) and 1q34 amplification. Chromosome 13q deletion, previously thought to predict poor outcome, is not a predictor following the use of newer agents. The ISS system incorporating the cytogenetic changes (Revised ISS) is the most widely used method for assessing prognosis (Table 111-3). Microarray profiling has formed the basis for RNA-based prognostic staging systems. Genome sequencing efforts have allowed for characterization of critical genes, pathways, and clonal heterogeneity in myeloma. The median number of mutations per transcribed genome in myeloma is ~ 58 , and within the whole genome, it is >7000 . A very heterogeneous mutational landscape with no unifying mutation has been observed. The most frequently mutated genes are *KRAS* and *NRAS* ($\sim 20\%$ each), followed by *TP53*, *DLS3*, *FAM46C*, and *BRAF*, all mutated in 5–10% of patients. All other mutations were observed in $<5\%$ of the patients. These results are now being applied to develop new targeted personalized therapies in myeloma.

TREATMENT

Multiple Myeloma

MGUS, SMM, AND SOLITARY PLASMACYTOMA

No specific intervention is indicated for patients with MGUS. Follow-up once a year or less frequently is adequate except in higher-risk MGUS, where serum protein electrophoresis, complete blood count, creatinine, and calcium should be repeated every 6 months. A patient with MGUS and severe polyneuropathy is considered for therapeutic intervention if a causal relationship can be

TABLE 111-3 Risk Stratification in Myeloma

CHROMOSOMAL ABNORMALITIES (CA)		
METHOD	STANDARD RISK (80%) (EXPECTED SURVIVAL 6–7+ YEARS)	HIGH RISK (20%) (EXPECTED SURVIVAL 2–3 YEARS)
Karyotype	No chromosomal aberration	Any abnormality on conventional karyotype
FISH	t(11;14) del(13)	del(17p) t(4;14) t(14;16) t(14;20) amp 1q34
INTERNATIONAL STAGING SYSTEM (ISS)		
	STAGE	MEDIAN SURVIVAL, MONTHS
$\beta_2\text{M} < 3.5$, alb ≥ 3.5	I (28%) ^a	62
$\beta_2\text{M} < 3.5$, alb < 3.5 or $\beta_2\text{M} = 3.5\text{--}5.5$	II (39%)	44
$\beta_2\text{M} > 5.5$	III (33%)	29
REVISED INTERNATIONAL STAGING SYSTEM (R-ISS)		
Stage I: ISS stage 1; standard risk for CA and normal LDH		
Stage II: Patients not meeting criteria for stage I or stage III		
Stage III: ISS stage III and either high risk for CA or high LDH		
Other features suggesting high-risk disease:		
De novo plasma cell leukemia		
Extramedullary disease		
Elevated LDH		
High-risk gene expression profile		

^aPercentage of patients presenting at each stage.

Abbreviations: $\beta_2\text{M}$, serum β_2 -microglobulin in mg/L; alb, serum albumin in g/dL; FISH, fluorescent in situ hybridization; LDH, lactate dehydrogenase.

assumed, especially in the absence of any other potential causes for neuropathy. Therapy can include plasmapheresis and occasionally rituximab in patients with IgM MGUS or myeloma-like therapy in those with IgG or IgA disease. A subset of patients with MGUS develop renal dysfunction usually based on renal damage from the monoclonal antibody. The damage may affect the glomeruli, tubules, or vessels. No consensus exists on management, but lowering the level of the monoclonal antibody with bortezomib has had some advocates.

About 10% of patients have SMM and will have an indolent course demonstrating only slow progression of disease over many years. For patients with SMM, no specific therapeutic intervention is indicated, although early intervention with lenalidomide and dexamethasone may prevent progression from high-risk SMM to active MM. At present, patients with SMM only require antitumor therapy when myeloma-defining events are identified.

Patients with solitary bone plasmacytomas and extramedullary plasmacytomas may be expected to enjoy prolonged disease-free survival after local radiation therapy at a dose of ~40 Gy. Occult marrow involvement may occur at low incidence in patients with solitary bone plasmacytoma. Such patients are usually identified because their serum M component falls slowly or disappears initially after local therapy, only to return after a few months. These patients respond well to systemic therapy.

SYMPOMATIC MM

Patients with symptomatic myeloma require therapeutic intervention. In general, such therapy has two purposes: (1) systemic therapy to control myeloma; and (2) supportive care to control symptoms of the disease, its complications, and adverse effects of therapy. Therapy can significantly prolong survival and improve the quality of life for myeloma patients.

The therapy of myeloma includes an initial induction regimen followed by consolidation and/or maintenance therapy and, on subsequent progression, management of relapsed disease. All agents available for use at various stages of the therapy and their doses, schedules, and combinations are detailed in **Table 111-4**. Therapy is partly dictated by the patient's age and comorbidities, which may affect a patient's ability to undergo high-dose therapy and transplantation (**Fig. 111-6**).

Three important classes of agents approved for treatment of newly diagnosed MM are immunomodulatory agents, proteasome inhibitors, and targeted antibodies. Thalidomide, when combined with dexamethasone, achieved responses in two-thirds of newly diagnosed MM patients. Subsequently, lenalidomide, an immunomodulatory derivative of thalidomide, and bortezomib, a proteasome inhibitor, have each been combined with dexamethasone with high response rates (>80%) in newly diagnosed patients with MM. Importantly, their lower toxicity profile with improved efficacy has made them the preferred agents for induction therapy. Efforts to improve the depth and frequency of response have involved using three-drug regimens. The combination of lenalidomide with a proteasome inhibitor (bortezomib or carfilzomib) and dexamethasone achieves close to a 100% response rate and a >30% complete response (CR) rate, making this combination one of the preferred induction regimens in transplant-eligible patients. Other similar three-drug combinations (bortezomib, thalidomide, and dexamethasone or bortezomib, cyclophosphamide, and dexamethasone) also achieve >90% response rate. Addition of a fourth agent, daratumumab, an anti-CD38 antibody, is providing even deeper responses. Usually between four and six cycles of these combination regimens are utilized to achieve initial deep cytoreduction before consideration of high-dose therapy with autologous stem cell transplantation.

In patients who are not transplant candidates due to physiologic age >70 years, significant cardiopulmonary problems, or other comorbid illnesses, the same two- or three-drug combinations described above are considered standard of care as induction therapy with age- and frailty-guided dose and schedule modifications. Modified lenalidomide-bortezomib-dexamethasone (RVD lite) combination achieves high overall response rate (86%) and CR (32%). Intermittent pulses of melphalan, an alkylating agent, with prednisone (MP) are combined with novel agents to achieve superior response and survival outcomes. In patients >65 years old, combining thalidomide with MP (MPT) obtains higher response rates and overall survival compared with MP alone. Similarly, significantly improved response (71 vs 35%) and overall survival (3-year survival 72 vs 59%) were observed with the combination of bortezomib and MP compared with MP alone. Continuous use of the lenalidomide and dexamethasone combination appears to be superior to the MPT regimen, and its combination with the anti-CD38 antibody daratumumab provides even higher overall response (92.9%) and CR rates (46.7%) and improved survival; the combination of lenalidomide, dexamethasone, and daratumumab is a standard-of-care regimen for older adults with myeloma.

HIGH-DOSE THERAPY WITH AUTOLOGOUS STEM CELL TRANSPLANTATION

High-dose therapy (HDT) and consolidation/maintenance are standard practice in the majority of eligible patients. In patients who are transplant candidates, alkylating agents such as melphalan should be avoided because they damage stem cells and compromise the ability to collect stem cells. Similarly, in patients receiving lenalidomide, stem cells should be collected within 6 months because the continued use of lenalidomide may compromise the ability to collect adequate numbers of stem cells. Randomized studies comparing standard-dose therapy to high-dose melphalan therapy with hematopoietic stem cell support have shown that HDT can achieve higher overall response rates, with up to 25–40% additional CRs and prolonged progression-free and overall survival; however, few, if any, patients are cured. Although two successive HDTs (tandem

TABLE 111-4 Standard Therapeutic Agents in Myeloma

CLASS	AGENT	STANDARD DOSAGE AND ADMINISTRATION	COMBINATION	MYELOMA INDICATION
Immunomodulatory drugs (IMiD)	Thalidomide (T)	Oral 50–200 mg qd	TD, VTD	Newly diagnosed and relapsed
	Lenalidomide (R)	Oral 5–25 mg daily × 21 days q 4 weeks	RD, RVD, DaRD, ERD, KRD, IRD	Newly diagnosed, maintenance, and relapsed
	Pomalidomide (P)	Oral 2–4 mg daily × 21 days q 4 weeks	PD	Relapsed
Proteasome inhibitors (PI)	Bortezomib (V)	IV or SC 1.3 mg/m ² days 1, 4, 8, 11 OR days 1, 8, 15	VD, VTD, VRD, DaVD, VCD	Newly diagnosed and relapsed
	Carfilzomib (K)	IV 20–56 mg/m ² days 1, 2, 8, 9, 15, 16 q 4 weeks	KD, KRD, KPD, Da KD, Da KRD, IsaKD	Newly diagnosed and relapsed
	Ixazomib (I)	Oral 4 mg days 1, 8, 15	IRD	Relapsed
Antibodies	Daratumumab (Da)	IV 16 mg/kg per week for 8 weeks then every 2 weeks for 16 weeks and then every 4 weeks thereafter	Dara, DaRD, DaVD, DaPD, DaKD	Newly diagnosed, maintenance, and relapsed
	Elotuzumab (E)	IV 10 mg/kg days 1, 8, 15, and 22 for first two cycles, then on days 1 and 15; along with RD	ERD, EPD	Relapsed
	Isatuximab (Isa)	IV 10 mg/kg weekly for 4 weeks and then every 2 weeks	IsaPD, IsaKD	Relapsed
	Belantamab mafodotin	IV 2.5 mg/kg once every 3 weeks		Relapsed or refractory - 4 prior lines of therapy
Selective inhibitor of nuclear export (SINE)	Selinexor (S)	Oral 80 mg on days 1 and 3 of each week	SVD	Relapsed
Histone deacetylase inhibitor	Panobinostat (Pa)	Oral 20 mg once every other day for 3 doses/week for 2 weeks every 21 days	PaVD	Relapsed
Alkylating agents	Melphalan (M)	Oral 0.25 mg/kg per day for 4 days (with P) every 4–6 weeks	MP, MPT, MPR, MPV, DaMPV, high-dose M	Newly diagnosed and relapsed conditioning
	Cyclophosphamide (C)	IV—300–500 mg/m ² weekly × 2 q 4 weeks Oral—50 mg qd × 21 days	VCD	Newly diagnosed and relapsed
	Bendamustine (B)	IV 70–90 mg days 1, 2 OR days 1, 8 q 4 weeks	BD or BVD	Relapsed
	MeIflufen (Me)	IV 40 mg day 1 (with D 40 mg on days 1, 8, 15, and 22) q 28 days	MeD	Relapsed or refractory - 4 prior lines of therapy
Cellular therapy	Idecabtagene vicleucel (Ide-cel)	IV 450 × 10 ⁶ cells	None	Relapsed or refractory - 4 prior lines of therapy with prior exposure to PI, IMiD, and anti-CD38 antibody
Glucocorticoid	Dexamethasone (D) Prednisone (P)	Oral 10–40 mg q week Oral 1 mg/kg		All stages

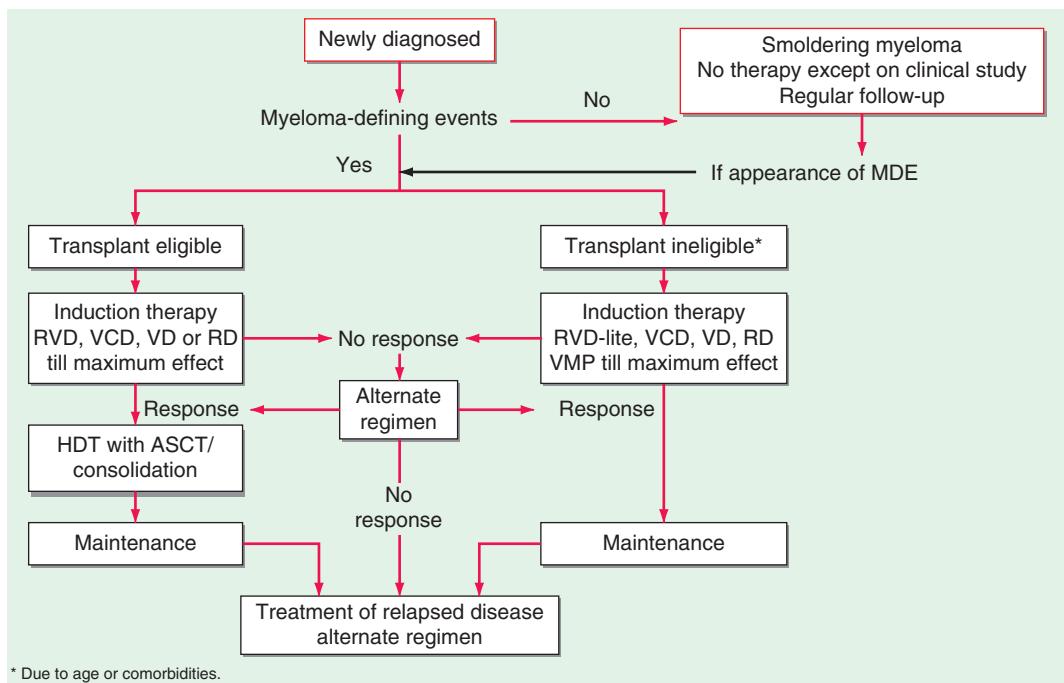


FIGURE 111-6 Treatment algorithm for multiple myeloma. C, cyclophosphamide; D, dexamethasone; M, melphalan; P, prednisone; R, lenalidomide; RVD-lite, weekly regimen; V, bortezomib. *Alternate regimen* indicates combinations including daratumumab, elotuzumab, panobinostat, carfilzomib, ixazomib, pomalidomide, or other agents. ASCT, autologous stem cell transplantation; HDT, high-dose therapy; MDE, myeloma-defining events.

transplantations) are more effective than single HDT, the benefit is only observed in the subset of patients who do not achieve a complete or very good partial response to the first transplantation, which is a rare subset. Moreover, a randomized study failed to show any significant difference in overall survival between early transplantation after induction therapy versus delayed transplantation at relapse. These data allow an option to delay transplantation, especially with the availability of newer agents and combinations. Allogeneic transplantations may also produce high response rates, but with significant toxicities. Nonmyeloablative allogeneic transplantation can reduce toxicity but is recommended only under the auspices of a clinical trial to exploit an immune graft-versus-myeloma effect while avoiding attendant toxicity.

Maintenance therapy prolongs remissions following standard-dose regimens as well as HDT. Two phase 3 studies have demonstrated improved progression-free survival, and one study showed prolonged overall survival in patients receiving lenalidomide compared to placebo as maintenance therapy after HDT. In non-transplant candidates, two phase 3 studies showed prolonged progression-free survival with lenalidomide maintenance after MP plus lenalidomide or lenalidomide plus dexamethasone induction therapy. Although concern arises regarding an increased incidence of second primary malignancies in patients receiving lenalidomide maintenance, its benefits in reducing the risk of progressive disease and death from myeloma far outweigh the small increased risk of second cancers. In patients with high-risk cytogenetics, lenalidomide and bortezomib or an oral proteasome inhibitor, ixazomib, show promise as maintenance combination therapy after transplantation.

RELAPSED DISEASE

Relapsed myeloma can be treated with a number of agents including lenalidomide and/or bortezomib, if previously not used. The second-generation proteasome inhibitor carfilzomib and immunomodulatory agent pomalidomide have shown efficacy in relapsed and refractory MM, even MM refractory to lenalidomide and bortezomib. An oral proteasome inhibitor, ixazomib, has also been approved in combination with lenalidomide and dexamethasone as an all-oral regimen for relapsed MM. Three antibodies are approved for treatment of relapsed MM. Daratumumab targeting CD38 achieves high response rates and improved progression-free survival as a single agent with further improvement in response and survival when added to bortezomib and dexamethasone or lenalidomide and dexamethasone. A formulation of daratumumab for subcutaneous administration provides decreased toxicity and improved convenience. Isatuximab, another antibody targeting CD38, achieves high response rates and improved progression-free survival in combination with pomalidomide or carfilzomib and dexamethasone. Elotuzumab, which targets SLAMF7, has shown significant activity in combination with lenalidomide and dexamethasone in relapsed/refractory myeloma but not as a single agent. Panobinostat, a histone deacetylase inhibitor, in combination with bortezomib and dexamethasone has been approved for treatment of relapsed refractory myeloma based on superior response and progression-free survival compared to bortezomib and dexamethasone alone. Two additional newer agents have unique mechanisms of action: selinexor is a first-in-class exportin inhibitor that blocks export of proteins from the cell nucleus, and mel氟ufen is an alkylating agent conjugated to a peptide to improve specific delivery to myeloma cells that express aminopeptidase required for cleaving of the peptide to deliver the drug intracellularly in myeloma cells. Both agents have been approved based on their effectiveness in relapsed/refractory myeloma. Another therapeutic focus has been to target BCMA, which is exclusively expressed on normal plasma cells and myeloma cells. An anti-BCMA antibody-drug conjugate, belantamab, targets BCMA and delivers auristatin to the tumor cells and achieves responses in relapsed/refractory myeloma. The drug has a unique ophthalmologic toxicity that requires close monitoring. Finally, a cellular therapy approved for myeloma is an anti-BCMA CAR transduced

T cell (idecogabtagene vicleucel [Idec-cel]), which is approved beyond fourth-line therapy. In patients with advanced myeloma with a median of six prior lines of treatment, 73% of patients receiving Idec-cel responded, and a CR rate of 33% was observed. Cytokine release syndrome and neurotoxicity remain primary toxicities requiring close monitoring and aggressive management. BCMA is also the target for a number of investigational agents including other CAR-T cell approaches as well as bispecific antibodies combining anti-BCMA with anti-CD3 antibody. Incorporation of the large number of active agents at various stages of treatment, including in newly diagnosed patients, is improving survival as well as quality of life.

THERAPY ENDPOINT

Improvement in the serum M component may lag behind the symptomatic improvement due to longer serum half-life (~3 weeks) of the immunoglobulin. The fall in M component depends on the rate of tumor kill and the fractional catabolic rate of immunoglobulin. Serum and urine light chains with a functional half-life of ~6 h may fall much quicker within the first week of treatment. Because urine light chain levels may relate to renal tubular function, they are not a reliable measure of tumor cell kill in patients with renal dysfunction. Achieving CR, defined as disappearance of serum and urine monoclonal protein with normal bone marrow by light microscopy, has been a standard goal of therapy. However, sequencing or multicolor flow cytometry-based assessment of minimal residual disease (MRD) in bone marrow to measure the presence of one myeloma cell in a million cells is being considered as an important new endpoint, especially in newly diagnosed patients. Absence of MRD at this sensitivity predicts for both longer progression-free survival and longer overall survival. Although patients may not achieve complete remission, clinical responses may last for long periods of time in small numbers of patients.

The median overall survival of patients with myeloma is 8+ years, with subsets of younger patients surviving >10 years. The major causes of death are progressive myeloma, renal failure, sepsis, or therapy-related myelodysplasia. Nearly a quarter of patients die of myocardial infarction, chronic lung disease, diabetes, or stroke, which are all intercurrent illnesses related more to the age of the patient group than to the tumor.

SUPPORTIVE THERAPY

Herpes zoster prophylaxis is indicated if bortezomib is used, and neuropathy attendant to bortezomib can be decreased both by its subcutaneous administration and by administration on a weekly schedule. Lenalidomide use requires prophylaxis for deep-vein thrombosis (DVT) with either aspirin or, if patients are at a greater risk of DVT, warfarin, low-molecular-weight heparin, or direct-acting anticoagulants. Patients receiving anti-BCMA CAR-T cell therapy may need supplementation with intravenous γ globulin due to induction of prolonged hypogammaglobulinemia.

Supportive care directed at the anticipated complications of the disease may be as important as primary antitumor therapy. Hypercalcemia generally responds well to bisphosphonates, glucocorticoid therapy, hydration, and natriuresis and rarely requires calcitonin as well. Bisphosphonates (e.g., pamidronate 90 mg or zoledronate 4 mg initially once a month for 12–24 months and later every 2–3 months) reduce osteoclastic bone resorption and preserve performance status and quality of life, decrease bone-related complications, and may also have antitumor effects. Osteonecrosis of the jaw and renal dysfunction can occur in a minority of patients receiving bisphosphonate therapy. Denosumab is an alternative agent administered intravenously at 120 mg monthly and achieves a similar level of effect as bisphosphonates to prevent bone-related complications in myeloma. Treatments aimed at strengthening the skeleton such as fluorides, calcium, and vitamin D, with or without androgens, have been suggested but are not of proven efficacy. Kyphoplasty or vertebroplasty should be considered in patients with painful collapsed vertebra. Iatrogenic worsening of renal function

may be prevented by maintaining a high fluid intake to prevent dehydration and enhance excretion of light chains and calcium. In the event of acute renal failure, plasmapheresis is ~10 times more effective at clearing light chains than peritoneal dialysis; however, its role in reversing renal failure remains controversial. Importantly, reducing the protein load by effective antitumor therapy with agents such as bortezomib may result in improvement in renal function in over half of the patients. Use of lenalidomide in renal failure is possible but requires dose modification because it is renally excreted. Urinary tract infections should be watched for and treated early. Plasmapheresis may be the treatment of choice for hyperviscosity syndromes. Although the pneumococcus is a dreaded pathogen in myeloma patients, pneumococcal polysaccharide vaccines may not elicit an antibody response. The pneumococcal conjugate vaccines are more protective. Prophylactic administration of intravenous γ globulin preparations is used in the setting of recurrent serious infections. Chronic oral antibiotic prophylaxis is not warranted. Patients developing neurologic symptoms in the lower extremities, severe localized back pain, or problems with bowel and bladder control may need emergency MRI and local radiation therapy and glucocorticoids if cord compression is identified. In patients in whom neurologic deficit is increasing or substantial, emergent surgical decompression may be necessary. Most bone lesions respond to analgesics and systemic therapy, but certain painful lesions may respond more promptly to localized radiation. The anemia associated with myeloma may respond to erythropoietin along with hematins (iron, folate, cobalamin). The pathogenesis of the anemia should be established and specific therapy instituted, whenever possible.

WALDENSTRÖM'S MACROGLOBULINEMIA

In 1948, Waldenström described a malignancy of lymphoplasmacytoid cells that secreted IgM. In contrast to myeloma, the disease was associated with lymphadenopathy and hepatosplenomegaly, but the major clinical manifestation was hyperviscosity syndrome. The disease resembles the related diseases CLL, myeloma, and lymphocytic lymphoma. It originates from a post-germinal center B cell that has undergone somatic mutations and antigenic selection in the lymphoid follicle and has the characteristics of an IgM-bearing memory B cell. Waldenström's macroglobulinemia (WM) and IgM myeloma follow a similar clinical course, but therapeutic options are different. The diagnosis of IgM myeloma is usually reserved for patients with lytic bone lesions and predominant infiltration with CD138+ plasma cells in the bone marrow. Such patients are at greater risk of pathologic fractures than patients with WM.

A familial occurrence is common in WM, but its molecular bases are yet unclear. A distinct *MYD88* L265P somatic mutation is present in >90% of patients with WM and the majority of IgM MGUS. Other commonly occurring mutations include *CXCR4* (30–40%), *ARID1A* (17%), and *CD79B* (8–15%). Presence of *MYD88* mutation status is now used as a diagnostic test to discriminate WM from marginal zone lymphomas (MZLs), IgM-secreting myeloma, and CLL with plasmacytic differentiation. This mutation also explains the molecular pathogenesis of the disease with involvement of Toll-like receptor (TLR) and interleukin 1 receptor (IL-1R) signaling leading to activation of IL-1R-associated kinase (IRAK) 4 and IRAK1 followed by nuclear factor- κ B (NF- κ B) activation. *MYD88* mutation also triggers Bruton's tyrosine kinase (BTK) and hemopoietic cell kinase (HCK)-mediated growth and survival signaling, which are now important therapeutic targets in WM. *CXCR4* mutations induce AKT and extracellular regulated kinase 1/2 (ERK1/2) signaling. This pathway can lead to development of drug resistance in the presence of its ligand CXCL12.

The disease is similar to myeloma in being slightly more common in men and occurring with increased incidence with increasing age (median age 64 years). The IgM in some patients with macroglobulinemia may have specificity for myelin-associated glycoprotein (MAG), a protein that has been associated with demyelinating disease of the peripheral nervous system and may be lost earlier and to a

greater extent than the better-known myelin basic protein in patients with multiple sclerosis. Sometimes patients with macroglobulinemia develop a peripheral neuropathy, and half of these patients are positive for anti-MAG antibody. The neuropathy may precede the appearance of the neoplasm. The whole process may begin with a viral infection that may elicit an antibody response that cross-reacts with a normal tissue component.

Like myeloma, the disease involves the bone marrow, but unlike myeloma, it does not cause bone lesions or hypercalcemia. Bone marrow shows >10% infiltration with lymphoplasmacytic cells (surface IgM+, CD19+, CD20+, and CD22+, rarely CD5+, but CD10– and CD23–) with an increase in number of mast cells. Like myeloma, an M component is present in the serum in excess of 30 g/L (3 g/dL), but unlike myeloma, the size of the IgM paraprotein results in little renal excretion, and only ~20% of patients excrete light chains. Therefore, renal disease is not common. The light chain isotype is kappa in 80% of the cases. Patients present with weakness, fatigue, and recurrent infections similar to myeloma patients, but epistaxis, visual disturbances, and neurologic symptoms such as peripheral neuropathy, dizziness, headache, and transient paresis are much more common in macroglobulinemia. Presence of *MYD88* and *CXCR4* mutations also affects disease presentation. Presence of *CXCR4* mutations is associated with higher bone marrow disease burden and higher incidence of hyperviscosity. Patients with wild-type *MYD88* show lower bone marrow disease burden.

Physical examination reveals adenopathy and hepatosplenomegaly, and ophthalmoscopic examination may reveal vascular segmentation and dilation of the retinal veins characteristic of hyperviscosity states. Patients may have a normocytic, normochromic anemia, but rouleaux formation and a positive Coombs test are much more common than in myeloma. Malignant lymphocytes are usually present in the peripheral blood. About 10% of macroglobulins are cryoglobulins. These are pure M components and are not the mixed cryoglobulins seen in rheumatoid arthritis and other autoimmune diseases. Mixed cryoglobulins are composed of IgM or IgA complexed with IgG, for which they are specific. In both cases, Raynaud's phenomenon and serious vascular symptoms precipitated by the cold may occur, but mixed cryoglobulins are not commonly associated with malignancy. Patients suspected of having a cryoglobulin based on history and physical examination should have their blood drawn into a warm syringe and delivered to the laboratory in a container of warm water to avoid errors in quantitating the cryoglobulin.

TREATMENT

Waldenström's Macroglobulinemia

A diagnosis of WM requires lymphoplasmacytic infiltrate of any level in the bone marrow and an IgM monoclonal paraprotein of any size. Treatment is usually not initiated unless the disease is symptomatic or increasing anemia, hyperviscosity, lymphadenopathy, or hepatosplenomegaly is present. Control of serious hyperviscosity symptoms such as an altered state of consciousness or paresis can be achieved acutely by plasmapheresis because 80% of the IgM paraprotein is intravascular. The median survival of affected individuals is ~50 months. However, many patients with WM have indolent disease that does not require therapy. Pretreatment parameters including older age, male sex, general symptoms, and cytopenias define a high-risk population. BTK inhibitors (ibrutinib), alkylating drugs (bendamustine and cyclophosphamide), and proteasome inhibitors (bortezomib, carfilzomib, and ixazomib), alone or more frequently in combination with rituximab, are considered as first-line therapy for symptomatic patients with WM. Ibrutinib targets the constitutively activated BTK. In patients with one prior line of therapy, the overall response to ibrutinib was 91%. Best responses to ibrutinib are observed in patients with mutated *MYD88* and wild-type *CXCR4* status, while delayed and lower response rates to ibrutinib are observed in patients with mutated *CXCR4*. At first relapse, in patients with an initial durable response,

either the previous regimen or another primary therapy regimen can be used. The therapeutic choice is dependent upon the genomic features, drug availability, and the patient's clinical profile.

Rituximab can produce an IgM flare, so either plasmapheresis should be used before rituximab or its use should be initially withheld in patients with high IgM levels. Fludarabine (25 mg/m² per d for 5 days every 4 weeks) and cladribine (0.1 mg/kg per d for 7 days every 4 weeks) are also highly effective single agents. With identification of the *MYD88* mutation, novel BTK inhibitors (acalabrutinib, zanubrutinib, and tirabrutinib), inhibitors targeting IRAK1/4, and the BCL2 antagonist venetoclax are being explored for the treatment of WM. Although HDT plus autologous transplantation is an option, its use has declined due to the availability of other effective agents.

POEMS SYNDROME

The features of this syndrome are polyneuropathy, organomegaly, endocrinopathy, *M*-protein, and skin changes (POEMS). Diagnostic criteria are described in Table 111-1. Patients usually have a severe, progressive sensorimotor polyneuropathy associated with sclerotic bone lesions from myeloma. Polyneuropathy occurs in ~1.4% of myelomas, but the POEMS syndrome is only a rare subset of that group. Unlike typical myeloma, hepatomegaly and lymphadenopathy occur in about two-thirds of patients, and splenomegaly is seen in one-third. The lymphadenopathy frequently resembles Castleman's disease histologically, a condition that has been linked to IL-6 overproduction. The endocrine manifestations include amenorrhea in women and impotence and gynecomastia in men. Hyperprolactinemia due to loss of normal inhibitory control by the hypothalamus may be associated with other central nervous system manifestations such as papilledema and elevated cerebrospinal fluid pressure and protein. Type 2 diabetes mellitus occurs in about one-third of patients. Hypothyroidism and adrenal insufficiency are occasionally noted. Skin changes are diverse: hyperpigmentation, hypertrichosis, skin thickening, and digital clubbing. Other manifestations include peripheral edema, ascites, pleural effusions, fever, and thrombocytosis. Not all the components of POEMS syndrome may be present initially.

The pathogenesis of the disease is unclear, but high circulating levels of the proinflammatory cytokines IL-1, IL-6, VEGF, and TNF have been documented, and levels of the inhibitory cytokine transforming growth factor β are lower than expected. Treatment of the myeloma may result in an improvement in the other disease manifestations.

Patients are often treated similarly to those with myeloma. Plasmapheresis does not appear to be of benefit in POEMS syndrome. Patients presenting with isolated sclerotic lesions may have resolution of neuropathic symptoms after local therapy for plasmacytoma with radiotherapy. Similar to MM, novel agents and HDT with autologous stem cell transplantation have been pursued in selected patients and have been associated with prolonged progression-free survival.

HEAVY CHAIN DISEASES

The heavy chain diseases are rare lymphoplasmacytic malignancies. Their clinical manifestations vary with the heavy chain isotype. Patients have absence of light chain and secrete a defective heavy chain that usually has an intact Fc fragment and a deletion in the Fd region. Gamma, alpha, and mu heavy chain diseases have been described, but no reports of delta or epsilon heavy chain diseases have appeared. Molecular biologic analysis of these tumors has revealed structural genetic defects that may account for the aberrant chain secreted.

GAMMA HEAVY CHAIN DISEASE (FRANKLIN'S DISEASE)

This disease affects individuals of widely different age groups and countries of origin. It is characterized by lymphadenopathy, fever, anemia, malaise, hepatosplenomegaly, and weakness. It is frequently associated with autoimmune diseases, especially rheumatoid arthritis. Its most distinctive symptom is palatal edema, resulting from involvement of nodes in Waldeyer's ring, and this may progress to produce

respiratory compromise. The diagnosis depends on the demonstration of an anomalous serum M component (often <20 g/L [<2 g/dL]) that reacts with anti-IgG but not anti-light chain reagents. The M component is typically present in both serum and urine. Most of the paraproteins have been of the γ_1 subclass, but other subclasses have been seen. The patients may have thrombocytopenia, eosinophilia, and nondiagnostic bone marrow that may show increased numbers of lymphocytes or plasma cells that do not stain for light chain. Patients usually have a rapid downhill course and die of infection; however, some patients have survived 5 years with chemotherapy. Therapy is indicated when symptomatic and involves chemotherapeutic combinations used in low-grade lymphoma. Rituximab has also been reported to show efficacy.

ALPHA HEAVY CHAIN DISEASE (SELIGMANN'S DISEASE)

This is the most common of the heavy chain diseases. It is closely related to a malignancy known as Mediterranean lymphoma, a disease that affects young persons in parts of the world where intestinal parasites are common, such as the Mediterranean, Asia, and South America. The disease is characterized by an infiltration of the lamina propria of the small intestine with lymphoplasmacytoid cells that secrete truncated alpha chains. Demonstrating alpha heavy chains is difficult because the alpha chains tend to polymerize and appear as a smear instead of a sharp peak on electrophoretic profiles. Despite the polymerization, hyperviscosity is not a common problem in alpha heavy chain disease. Without J chain-facilitated dimerization, viscosity does not increase dramatically. Light chains are absent from serum and urine. The patients present with chronic diarrhea, weight loss, and malabsorption and have extensive mesenteric and paraaortic adenopathy. Respiratory tract involvement occurs rarely. Patients may vary widely in their clinical course. Some may develop diffuse aggressive histologies of malignant lymphoma. Chemotherapy may produce long-term remissions. Rare patients appear to have responded to antibiotic therapy, raising the question of the etiologic role of antigenic stimulation, perhaps by some chronic intestinal infection. Chemotherapy plus antibiotics may be more effective than chemotherapy alone. IPSID is recognized as an infectious pathogen-associated human lymphoma associated with *Campylobacter jejuni*. It involves mainly the proximal small intestine, resulting in malabsorption, diarrhea, and abdominal pain. IPSID is associated with excessive plasma cell differentiation and produces truncated alpha heavy chain proteins lacking the light chains as well as the first constant domain. Early-stage IPSID responds to antibiotics (30–70% complete remission). Most untreated IPSID patients progress to lymphoplasmacytic and immunoblastic lymphoma. Patients not responding to antibiotic therapy are considered for treatment with combination chemotherapy used to treat low-grade lymphoma.

MU HEAVY CHAIN DISEASE

The secretion of isolated mu heavy chains into the serum appears to occur in a very rare subset of patients with CLL. The only features that may distinguish patients with mu heavy chain disease are the presence of vacuoles in the malignant lymphocytes and the excretion of kappa light chains in the urine. The diagnosis requires ultracentrifugation or gel filtration to confirm the nonreactivity of the paraprotein with the light chain reagents because some intact macroglobulins fail to interact with these serums. The tumor cells seem to have a defect in the assembly of light and heavy chains because they appear to contain both in their cytoplasm. Such patients are not treated differently from other patients with CLL (Chap. 107).

FURTHER READING

- ATTAL M et al: IFM 2009 study. Lenalidomide, bortezomib, and dexamethasone with transplantation for myeloma. *N Engl J Med* 376:1311, 2017.
- CORRE J et al: Risk factors in multiple myeloma: is it time for a revision? *Blood* 137:16, 2021.
- HIDESHIMA T, ANDERSON KC: Signaling pathway mediating myeloma cell growth and survival. *Cancers (Basel)* 13:216, 2021.

- HILLENGASS J et al: International myeloma working group consensus recommendations on imaging in monoclonal plasma cell disorders. *Lancet Oncol* 20:e302, 2019.
- HUNTER ZR et al: Genomics, signaling, and treatment of Waldenström macroglobulinemia. *J Clin Oncol* 35:994, 2017.
- KUMAR S et al: International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *Lancet Oncol* 17:e328, 2016.
- MOREAU P et al: Treatment of relapsed and refractory multiple myeloma: Recommendations from the International Myeloma Working Group. *Lancet Oncol*. 22:e105, 2021.
- RAJKUMAR SV et al: International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol* 15:e538, 2014.
- ROBIOU DU PONT S et al: Genomics of multiple myeloma. *J Clin Oncol* 35:963, 2017.
- TERPOS E et al: Treatment of multiple myeloma-related bone disease: Recommendations from the Bone Working Group of the International Myeloma Working Group. *Lancet Oncol* 22:e119, 2021.

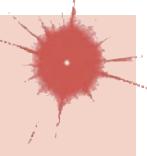
and to eliminate proteins that misfold. However, genetic mutation, incorrect processing, and other factors may favor misfolding, with consequent loss of normal protein function and intracellular or extracellular aggregation. Many diseases, ranging from cystic fibrosis to Alzheimer's disease, are now known to involve protein misfolding. In the amyloidoses, the aggregates are typically extracellular, and the misfolded protein subunits assume a common antiparallel, β -pleated sheet-rich structural conformation that leads to the formation of higher-order oligomers and then fibrils with unique staining properties. The term *amyloid* was coined around 1854 by the pathologist Rudolf Virchow, who thought that these deposits resembled starch (Latin *amylum*) under the microscope.

Amyloid diseases, defined by the biochemical nature of the protein composing the fibril deposits, are classified according to whether they are systemic or localized, whether they are acquired or inherited, and their clinical patterns (Table 112-1). The standard nomenclature is AX, where A indicates amyloidosis and X represents the protein present in the fibril. This chapter focuses primarily on the systemic forms. AL refers to amyloid composed of immunoglobulin light chains (LCs); this disorder, formerly termed *primary systemic amyloidosis*, arises from a clonal B-cell or plasma cell disorder and can be associated with myeloma or lymphoma. ATTR, the most prevalent of the *familial amyloidoses*, refers to amyloid derived from wild-type or mutated transthyretin (TTR), the transport protein for thyroid hormone and retinol-binding protein. AA amyloid is composed of the acute-phase reactant protein serum amyloid A (SAA) and occurs in the setting of chronic inflammatory or infectious diseases; for this reason, this type was formerly known as *secondary amyloidosis*. $A\beta_2M$ amyloid results from misfolded β_2 -microglobulin, occurring in individuals with long-standing renal disease who have undergone dialysis, typically for years. $A\beta$, the most common form of localized amyloidosis, is found in the brain of patients with Alzheimer's disease after abnormal proteolytic processing and aggregation of polypeptides derived from the amyloid precursor protein.

Diagnosis and treatment of the amyloidoses rest upon the histopathologic identification of amyloid deposits and immunohistochemical, biochemical, or genetic determination of amyloid type (Fig. 112-1). In the systemic amyloidoses, the clinically involved

112 Amyloidosis

John L. Berk, Vaishali Sanctorawala



GENERAL PRINCIPLES

Amyloidosis is the term for a group of protein misfolding disorders characterized by the extracellular deposition of insoluble polymeric protein fibrils in tissues and organs. A robust cellular machinery exists to chaperone proteins during the process of synthesis and secretion, to ensure that they achieve correct tertiary conformation and function,

TABLE 112-1 Amyloid Precursor Proteins and Their Clinical Syndromes

DESIGNATION	PRECURSOR	CLINICAL SYNDROME	CLINICAL INVOLVEMENT
Systemic Amyloidoses			
AL	Immunoglobulin light chain	Primary or myeloma-associated ^a	Any
AH	Immunoglobulin heavy chain	Rare variant of primary or myeloma-associated	Any
AA	Serum amyloid A protein	Secondary; reactive ^b	Renal, heart, other
$A\beta_2M$	β_2 -Microglobulin	Hemodialysis-associated	Synovial tissue, bone
ATTR	Transthyretin	Familial (mutant) Age-related (wild type)	Cardiac, peripheral and autonomic nerves, soft tissues, spine, bladder
AApoAI	Apolipoprotein AI	Familial	Hepatic, renal
AApoAI	Apolipoprotein AI	Familial	Renal
AGel	Gelsolin	Familial	Cornea, cranial nerves, skin, renal
AFib	Fibrinogen α I	Familial	Renal, vascular
ALys	Lysozyme	Familial	Renal, hepatic
ALECT2	Leukocyte chemotactic factor 2	Undefined	Renal
Localized Amyloidoses			
$A\beta$	Amyloid β protein	Alzheimer's disease; Down's syndrome	Central nervous system
ACys	Cystatin C	Cerebral amyloid angiopathy	Central nervous system, vascular
APrP	Prion protein	Spongiform encephalopathies	Central nervous system
AIAPP	Islet amyloid polypeptide (amylin)	Diabetes-associated	Pancreas
ACal	Calcitonin	Medullary carcinoma of the thyroid	Thyroid
AANF	Atrial natriuretic factor	Atrial fibrillation	Cardiac atria
APro	Prolactin	Endocrinopathy	Pituitary
ASgl	Semenogelin I	Age-related; incidental autopsy or biopsy finding	Seminal vesicles

^aLocalized AL deposits can occur in skin, conjunctiva, urinary bladder, and the tracheobronchial tree. ^bSecondary to chronic inflammation or infection or to a hereditary periodic fever syndrome such as familial Mediterranean fever.

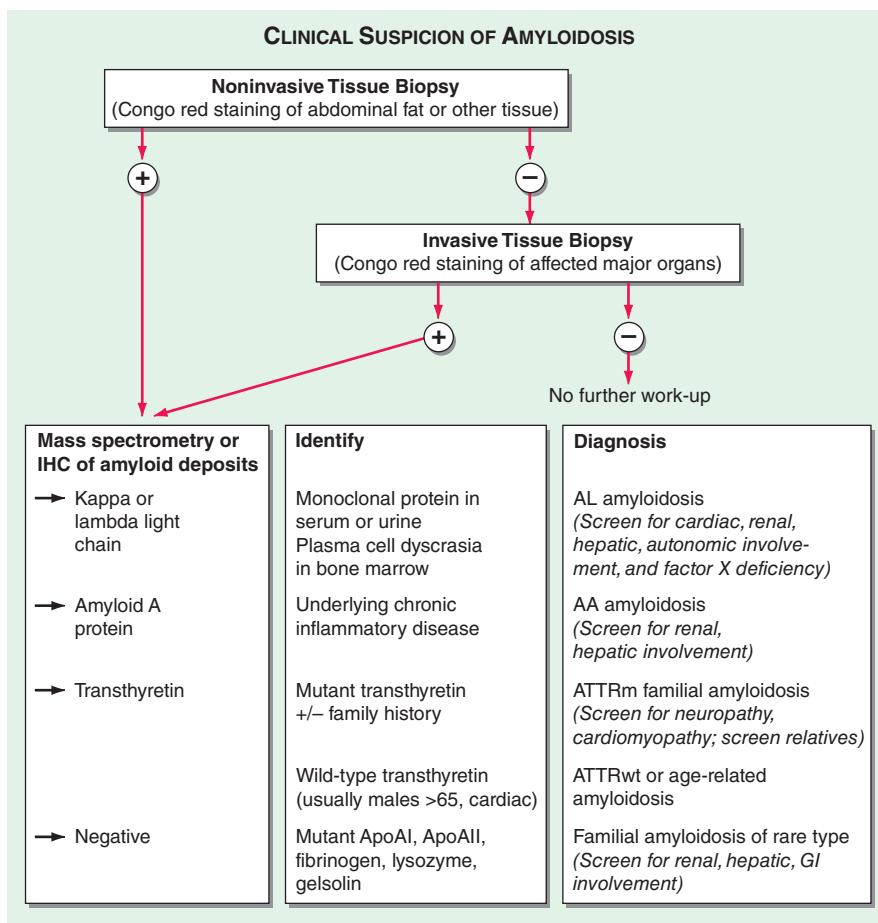


FIGURE 112-1 Algorithm for the diagnosis of amyloidosis and determination of type. Clinical suspicion: unexplained nephropathy, cardiomyopathy, neuropathy, enteropathy, arthropathy, and macroglossia. ApoAI, apolipoprotein AI; ApoAI, apolipoprotein AI; GI, gastrointestinal; IHC, immunohistochemistry.

organs can be biopsied, but amyloid deposits may be found in any tissue of the body. Historically, blood vessels of the gingiva or rectal mucosa were often examined, but the most easily accessible tissue—positive in more than 80% of patients with systemic amyloidosis—is abdominal fat. After local anesthesia, fat is aspirated with a 16-gauge needle from the subcutaneous layer of the abdominal wall. Fat globules expelled onto a glass slide can be stained for amyloid, thus avoiding a surgical procedure. If this material is negative, more invasive biopsies of the kidney, heart, liver, tongue, or gastrointestinal tract can be considered in patients in whom amyloidosis is suspected. The regular β -sheet structure of amyloid deposits exhibits a unique “apple green” birefringence by polarized light microscopy when stained with Congo red dye; other regular protein structures (e.g., collagen) appear white under these conditions. The 10-nm-diameter fibrils can also be visualized by electron microscopy of paraformaldehyde-fixed tissue. Once amyloid is found, the precursor protein type must be determined by immunohistochemistry, immunolectron microscopy, or extraction and biochemical analysis employing mass spectrometry; gene sequencing is used to identify mutants causing hereditary amyloidosis. The patient’s history, physical findings, and clinical presentation, including age and ethnic origin, organ system involvement, underlying diseases, and family history, may provide helpful clues as to the type of amyloidosis. However, there can be considerable overlap in clinical presentations, and accurate typing is essential to guide appropriate therapy.

The mechanisms of fibril formation and tissue toxicity remain controversial. The “amyloid hypothesis,” as it is currently understood, proposes that precursor proteins undergo a process of reversible unfolding or misfolding; misfolded proteins form oligomeric aggregates, higher-order polymers, and then fibrils that deposit in tissues. Accumulating evidence suggests that the oligomeric intermediates may constitute the most toxic species. Oligomers are more capable than large fibrils of interacting

with cells and inducing formation of reactive oxygen species and stress signaling. Ultimately, the fibrillar tissue deposits are likely to interfere with normal organ function. A more sophisticated understanding of the mechanisms leading to amyloid formation and cell and tissue dysfunction will continue to provide new targets for therapies.

The clinical syndromes of the amyloidoses are associated with relatively nonspecific alterations in routine laboratory tests. Blood counts are usually normal, although the erythrocyte sedimentation rate is frequently elevated. Patients with glomerular kidney involvement generally have proteinuria, often in the nephrotic range, leading to hypoalbuminemia that may be severe; patients with serum albumin levels <2 g/dL generally have pedal edema or anasarca. Amyloid cardiomyopathy is characterized by concentric ventricular hypertrophy and diastolic dysfunction associated with elevation of brain natriuretic peptide (BNP) or N-terminal pro-brain natriuretic peptide (NT-proBNP) as well as troponin. These cardiac biomarkers can be used for disease staging, prognostication, and disease activity monitoring in patients with AL amyloidosis. Notably, renal insufficiency can falsely elevate levels of these biomarkers. Recently, biomarkers of cardiac remodeling—that is, matrix metalloproteinases and tissue inhibitors of metalloproteinases—have been found to be altered in the serum of patients with amyloid cardiomyopathy. Electrocardiographic and echocardiographic features of amyloid cardiomyopathy are described below. Patients with liver involvement, even when advanced, usually develop cholestasis with an elevated alkaline phosphatase concentration with minimal alteration of the aminotransferases and preservation

of synthetic function. In AL amyloidosis, endocrine organs may be infiltrated with fibrils, and hypothyroidism, hypoadrenalinism, or even hypopituitarism can occur. Although none of these findings is specific for amyloidosis, the presence of abnormalities in multiple organ systems should raise suspicions of the diagnosis.

■ AL AMYLOIDOSIS

Etiology and Incidence AL amyloidosis is most frequently caused by a clonal expansion of bone marrow plasma cells that secrete a monoclonal immunoglobulin LC depositing as amyloid fibrils in tissues. Whether the clonal plasma cells produce a LC that misfolds and leads to AL amyloidosis or an LC that folds properly, allowing the cells to inexorably expand over time and develop into multiple myeloma (Chap. 111), may depend upon primary sequence of the clonal LC or other genetic or epigenetic factors. AL amyloidosis can occur with multiple myeloma or other B lymphoproliferative diseases, including non-Hodgkin’s lymphoma (Chap. 108) and Waldenström’s macroglobulinemia (Chap. 111). AL amyloidosis is the most common type of systemic amyloidosis diagnosed in North America. Its incidence has been estimated at 4.5 cases/100,000 population; however, ascertainment continues to be inadequate, and the true incidence may be much higher. AL amyloidosis, like other plasma cell disorders, usually occurs after age 40 and is often progressive and fatal if untreated.

Pathology and Clinical Features Amyloid deposits are usually widespread in AL amyloidosis and can be present in the interstitium of any organ outside the central nervous system. The amyloid fibril deposits are composed of full-length 23-kDa monoclonal immunoglobulin LCs as well as fragments. Accessory molecules co-deposited with LC fibrils (as well as with other amyloid fibrils) include serum amyloid P component, apolipoproteins e and A-IV, glycosaminoglycans, and

metal ions. Although all kappa and lambda LC subtypes have been identified in AL amyloid fibrils, lambda subtypes predominate. The lambda 6 subtype appears to have unique structural properties that predispose it to fibril formation, often in the kidney.

AL amyloidosis is often a rapidly progressive disease that presents as a pleiotropic set of clinical syndromes, recognition of which is key for initiation of the appropriate workup. Nonspecific symptoms of fatigue and weight loss are common; however, the diagnosis is rarely considered until symptoms referable to a specific organ develop. The kidneys are the most frequently involved organ and are affected in 70–80% of patients. Renal amyloidosis usually manifests as proteinuria, often in the nephrotic range and associated with hypoalbuminemia, secondary hypercholesterolemia and hypertriglyceridemia, and edema or anasarca. In some patients, interstitial rather than glomerular amyloid deposition can produce azotemia without proteinuria. The heart is the second most commonly affected organ (50–60% of patients), and cardiac involvement is the leading cause of death from AL amyloidosis. Early on, the electrocardiogram may show low voltage in the limb leads with a pseudo-infarct pattern. Echocardiographic features of disease include concentrically thickened ventricles and diastolic dysfunction with an abnormal global longitudinal strain pattern; a “sparkly” appearance has been described but is often not seen with modern high-resolution echocardiographic techniques. Poor atrial contractility occurs even in sinus rhythm, and patients with cardiac amyloidosis are at risk for development of atrial thrombi and stroke. Cardiac MRI can show increased wall thickness, and characteristic delayed enhancement of the subendocardium has been described following injection of gadolinium contrast. Nervous system symptoms include peripheral sensorimotor neuropathy and/or autonomic dysfunction manifesting as gastrointestinal motility disturbances (early satiety, diarrhea, constipation), dry eyes and mouth, impotence, orthostatic hypotension, and/or neurogenic bladder. Macroglossia (Fig. 112-2A), a pathognomonic sign of AL amyloidosis, is seen in only ~10% of patients. Liver involvement causes cholestasis and hepatomegaly. The spleen is frequently involved, and there may be functional hypersplenism in the absence of significant splenomegaly. Many patients experience “easy bruising” due to amyloid deposits in capillaries or deficiency of clotting factor X due to binding to amyloid fibrils; cutaneous ecchymoses appear, particularly around the eyes, producing another uncommon but pathognomonic finding, the “raccoon-eye” sign (Fig. 112-2B). Other findings include nail dystrophy (Fig. 112-2C), alopecia, and amyloid arthropathy with thickening of synovial membranes in the wrists and shoulders. The presence of a multisystemic illness or general fatigue along with any of these clinical syndromes should prompt a workup for amyloidosis.

Diagnosis Identification of an underlying clonal plasma cell or B lymphoproliferative process and a clonal LC are key to the diagnosis of AL amyloidosis. Serum protein electrophoresis and urine protein electrophoresis, although of value in multiple myeloma, are *not* useful screening tests if AL amyloidosis is suspected because the clonal LC or whole immunoglobulin often is not present in sufficient amounts to produce a monoclonal “M-spike” in the serum or LC (Bence Jones) protein in the urine. However, more than 90% of patients with AL amyloidosis have serum or urine monoclonal LC or whole immunoglobulin

detectable by immunofixation electrophoresis of serum (SIFE) or urine (UIFE) (Fig. 112-3A) or by nephelometric measurement of serum “free” LCs (i.e., LCs circulating in monomeric form rather than in an immunoglobulin tetramer with heavy chain). Examining the ratio as well as the absolute amount of serum-free LCs is essential, as renal insufficiency reduces LC clearance, nonspecifically elevating both isotypes. In addition, an increased percentage of plasma cells in the bone marrow—typically 5–30% of nucleated cells—is found in ~90% of patients. Kappa or lambda clonality should be demonstrated by flow cytometry, immunohistochemistry, or *in situ* hybridization for LC mRNA (Fig. 112-3B).

A monoclonal serum protein by itself is not diagnostic of amyloidosis, since monoclonal gammopathy of uncertain significance is common in older patients (Chap. 111). However, when monoclonal gammopathy of uncertain significance is found in patients with biopsy-proven amyloidosis, the AL type should be ruled out. Similarly, patients thought to have “smoldering myeloma” because of a modest elevation of bone-marrow plasma cells should be screened for AL amyloidosis if they have signs or symptoms of renal, cardiac, or neurologic disease. Accurate tissue amyloid typing is essential for appropriate treatment. Immunohistochemical staining of the amyloid deposits is useful if they selectively bind one LC antibody in preference to the other; some AL deposits bind antibodies nonspecifically. Immunoelectron microscopy is more reliable; laser capture microdissection and tandem mass spectrometry-based typing of the amyloid precursor protein has become the diagnostic standard. In ambiguous cases, other forms of amyloidosis should be thoroughly excluded with appropriate genetic and other testing.

Staging System and Risk Stratification The current staging systems for systemic AL amyloidosis are based on the biomarkers of plasma cell dyscrasia and cardiac and renal involvement. The Mayo 2004 staging system is based on the levels of NT-proBNP and cardiac troponins and was modified by European investigators to identify and classify very-high-risk patients. This cardiac staging system is the most widely used to determine patient management. This staging system was modified (Mayo 2012) to include clonal burden, assessed by dFLC (difference between involved and uninvolved circulating free light chain) concentration, which has independent ability to predict survival. Boston University investigators introduced a staging system incorporating BNP and troponin I that also is able to predict survival. Patients with AL amyloidosis with a very low (<50 mg/L) dFLC level have a significantly better outcome irrespective of cardiac stage. A renal staging system based on 24-h urine protein excretion and estimated glomerular filtration rate (eGFR) predicting the progression to dialysis at 2 years has also been developed and validated. Several other biomarkers have been shown to predict outcomes and survival but have not been incorporated in staging systems yet.

TREATMENT

AL Amyloidosis

Extensive multisystemic involvement typifies AL amyloidosis, and the median survival period without treatment is usually only ~1–2 years from the time of diagnosis. Current therapies target the



A



B



C

FIGURE 112-2 Clinical signs of AL amyloidosis. A. Macroglossia. **B.** Periorbital ecchymoses. **C.** Fingernail dystrophy.

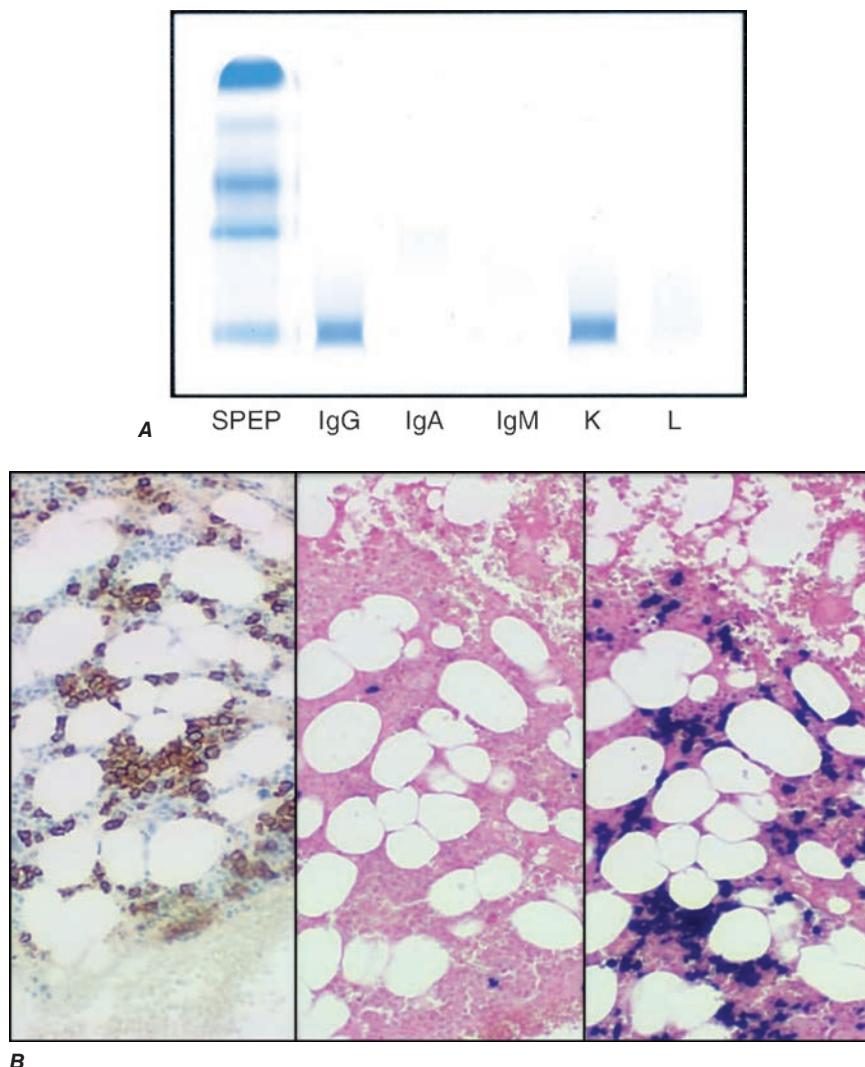


FIGURE 112-3 Laboratory features of AL amyloidosis. **A.** Serum immunofixation electrophoresis reveals an IgGκ monoclonal protein in this example; serum protein electrophoresis is often normal. **B.** Bone marrow biopsy sections stained by immunohistochemistry with antibody to CD138 (syndecan, highly expressed on plasma cells) (left) or by *in situ* hybridization with fluorescein-tagged probes (Ventana Medical Systems) binding to κ mRNA (center) and λ mRNA (right) in plasma cells. (Photomicrograph courtesy of C. O'Hara; with permission.)

clonal bone marrow plasma cells, using approaches employed for multiple myeloma. Treatment with oral melphalan and prednisone can decrease the plasma cell burden but rarely leads to complete hematologic remission, meaningful organ responses, or improved survival and is no longer widely used. The substitution of dexamethasone for prednisone produces a higher response rate and more durable remissions, although dexamethasone is not always well tolerated by patients with significant edema or cardiac disease. High-dose intravenous (IV) melphalan followed by autologous stem cell transplantation (HDM/SCT) produces complete hematologic responses in ~40% of treated patients, as determined by loss of clonal plasma cells in the bone marrow and disappearance of the amyloidogenic monoclonal LC, as determined by SIFE/UIFE and free LC quantitation. Six to 12 months after achieving a hematologic response, improvements in organ function and quality of life may occur. Hematologic responses appear to be more durable after HDM/SCT than in multiple myeloma, with remissions continuing in some patients beyond 15 years without additional treatment. Unfortunately, only ~20–30% of all AL amyloidosis patients are suitable for aggressive treatment, and even at specialized treatment centers, transplantation-related mortality rates are higher than those for other hematologic diseases because of impaired organ function at initial presentation. Amyloid cardiomyopathy, poor nutritional and performance status, and multiorgan disease contribute to excess morbidity and mortality. A bleeding diathesis

resulting from adsorption of clotting factor X to amyloid fibrils also increases mortality rates; however, this syndrome occurs in only 5–10% of patients. A randomized multicenter trial conducted in France compared oral melphalan and dexamethasone with HDM/SCT and failed to show a benefit of dose-intensive treatment, although the transplantation-related mortality rate in this study was very high. It has become clear that careful selection of patients and expert peritransplantation management are essential in reducing transplantation-related mortality.

For patients with AL amyloidosis and impaired cardiac function or arrhythmias due to involvement of the myocardium, the median survival period is only ~6 months without treatment. In these patients, cardiac transplantation can be performed and followed by HDM/SCT to eliminate the noxious LC clone and prevent amyloid deposition in the transplanted heart or other organs.

The best therapy for those who are transplant ineligible varies between centers and countries. A regimen of oral chemotherapy with melphalan and dexamethasone (MDex) had been the standard for patients not eligible for HDM/SCT for more than a decade. Regimens using bortezomib (a proteasome inhibitor) are now considered the standard of care in most patients with AL amyloidosis not eligible for SCT. There is a fine balance between chosen treatment regimens and toxicities, and patient characteristics should be considered when choosing a regimen; for example, treatment with bortezomib plus MDex can overcome the effects of both gain

of 1q21 (which confers a poorer outcome with oral melphalan) and t(11;14) (which confers a poorer outcome with bortezomib). Transplant-ineligible patients in whom bortezomib is contraindicated due to preexisting peripheral neuropathy can be treated with MDex or combinations based on immunomodulatory drugs (e.g., lenalidomide). High-risk patients represent ~20% of all individuals with AL amyloidosis and are a challenge owing to advanced cardiac stage (IIIb) or severe heart failure (New York Heart Association class III or IV).

Newer agents, such as the oral proteasome inhibitor ixazomib and the humanized anti-CD38 monoclonal antibody daratumumab, have also been evaluated in patients with relapsed or refractory disease. Anti-fibril small molecules and humanized monoclonal antibodies are also being tested. Clinical trials are essential in improving therapy for this rare disease.

Supportive care is important for patients with any type of amyloidosis. For nephrotic syndrome, diuretics and supportive stockings can ameliorate edema; angiotensin-converting enzyme inhibitors should be used with caution and have not been shown to slow renal disease progression. Effective diuresis can be facilitated with albumin infusions to raise intravascular oncotic pressure. Congestive heart failure due to amyloid cardiomyopathy is best treated with diuretics; it is important to note that digitalis, calcium channel blockers, and beta blockers are relatively contraindicated as they can interact with amyloid fibrils and produce heart block and worsening heart failure. Amiodarone has been used for atrial and ventricular arrhythmias. Automatic implantable defibrillators appear to have reduced effectiveness due to the thickened myocardium, but they may benefit some patients. Atrial ablation is an effective approach for atrial fibrillation. For conduction abnormalities, ventricular pacing may be indicated. Atrial contractile dysfunction is common in amyloid cardiomyopathy and associated with increased thromboembolic complications, prompting considerations of anti-coagulation even in the absence of atrial fibrillation. Autonomic neuropathy can be treated with α agonists such as midodrine to support postural blood pressure; gastrointestinal dysfunction may respond to motility or bulk agents. Nutritional supplementation, either oral or parenteral, is also important.

In localized AL amyloidosis, amyloid deposits can be produced by clonal plasma cells infiltrating local sites in the airways, bladder, skin, or lymph nodes (Table 112-1). These deposits may respond to surgical intervention or elimination of the responsible plasma cell clone by low-dose radiation therapy (typically only 20 Gy); systemic treatment generally is not appropriate. Patients should be referred to a center familiar with management of these rare manifestations of amyloidosis.

■ AA AMYLOIDOSIS

Etiology and Incidence AA amyloidosis can occur in association with almost any chronic inflammatory state (e.g., rheumatoid arthritis, inflammatory bowel disease, ankylosing spondylitis, familial Mediterranean fever [Chap. 369], or other periodic fever syndromes) or chronic infections such as tuberculosis, osteomyelitis, or subacute bacterial endocarditis. In the United States and Europe, AA amyloidosis has become less common, occurring in fewer than 2% of patients with these diseases, presumably because of advances in anti-inflammatory and antimicrobial therapies. It has also been described in association with Castleman's disease, lymphomas, and renal cell carcinoma, emphasizing the diagnostic importance of CT scanning to look for such tumors as well as serologic and microbiologic studies. In up to 30% of patients, AA amyloidosis can also be seen without any identifiable underlying disease. AA is the most frequent systemic amyloidosis that occurs in children.

Pathology and Clinical Features Organ involvement in AA amyloidosis usually begins in the kidneys. Hepatomegaly, splenomegaly, and autonomic neuropathy can also occur as the disease progresses; cardiomyopathy is a late manifestation in ~25% of patients. The symptoms and signs of AA disease cannot be reliably distinguished from

those of AL amyloidosis. AA amyloid fibrils are usually composed of an 8-kDa, 76-amino-acid N-terminal portion of the 12-kDa precursor protein SAA. This acute-phase apoprotein is synthesized in the liver and transported by high-density lipoprotein (HDL3) in the plasma. Several years of an underlying inflammatory disease causing chronic elevation of SAA levels usually precede fibril formation, although infections can lead to AA amyloid deposition more rapidly.

TREATMENT

AA Amyloidosis

Primary therapy for AA amyloidosis consists of treatment of the underlying inflammatory or infectious disease. Treatment that suppresses or eliminates the inflammatory state or infection decreases the SAA concentration, slowing the rate of amyloid fibril formation. For familial Mediterranean fever, colchicine at a dose of 1.2–1.8 mg/d is the standard treatment. However, colchicine has not been helpful for AA amyloidosis of other causes or for other amyloidoses. Tumor necrosis factor and interleukin 1 and interleukin 6 antagonists can effectively interrupt cytokine signaling that drives many inflammatory syndromes, inhibiting hepatic SAA production and limiting AA amyloid deposition. Development of a fibril-specific agent (eprodise) that interferes with the interaction of serum amyloid A protein and glycosaminoglycans to prevent or disrupt fibril formation failed in phase 3 trials.

■ ATTR AND AF AMYLOIDOSIS

The familial amyloidoses are autosomal dominant diseases in which mutated or variant plasma proteins misfold or aggregate to form beta-sheet rich amyloid deposits. These diseases are rare, with an estimated case incidence of <1/100,000 population in the United States, although founder effects in remote areas of Portugal, Sweden, and Japan produced a higher local prevalence of disease. The most prevalent form of hereditary amyloidosis arises from mutation of the abundant liver-derived plasma protein transthyretin (TTR, also known as *prealbumin*) and is termed hATTR amyloid. More than 130 TTR mutations typically conferring one-amino-acid substitutions have been described, with most inducing clinical ATTR amyloid disease. Toxic TTR oligomers and ATTR amyloid deposits target peripheral and autonomic nervous systems and the heart. One TTR variant, V122I, occurs in nearly 4% of the African-American and Afro-Caribbean populations and is associated with late-onset cardiac amyloidosis. The actual incidence and penetrance of disease in the African-American population is the subject of ongoing research, but considerations of V122I ATTR amyloidosis is warranted in African-American patients who present with concentric cardiac hypertrophy and evidence of diastolic heart failure, particularly in the absence of a history of hypertension or valvular disease. Other familial amyloidoses, caused by variant apolipoproteins AI or AII, gelsolin, fibrinogen Aα, or lysozyme, are reported with lower prevalence worldwide. New amyloidogenic serum proteins continue to be identified periodically, including leukocyte chemotactic factor LECT2, which is a cause of renal amyloidosis in Hispanic and Pakistani populations. Although the clustering of ALECT2 cases suggests heritability, no LECT2 gene-coding sequence variations have been identified.

Normal (wild-type) transthyretin can also misfold and aggregate to form ATTR amyloid, typically expressed in men beginning in the seventh decade with increasing prevalence with age. Formerly termed senile systemic amyloidosis, ATTRwt amyloid is reported at autopsy in 25% of hearts from patients who are 80 years and older. Although it is unclear why a wild-type protein becomes amyloidogenic, aging inefficiencies of intracellular quality-assurance mechanisms (termed the unfolded protein response) likely predispose to secretion of proteins prone to misaggregation. Due to the numbers of aging men globally, ATTRwt is the most prevalent and rapidly growing form of amyloidosis in the world today.

Clinical Features and Diagnosis hATTR amyloidosis has a varied presentation predicted by the specific TTR mutation. Consequently,

kindreds typically express similar disease timing and clinical course. Apparent sporadic presentations (no recognized family history) often reflect incomplete penetrance of the TTR mutation and not a spontaneous event. hATTR amyloidosis presents as familial amyloidotic polyneuropathy (nerve damage) or familial amyloidotic cardiomyopathy (heart damage). Peripheral neuropathy begins as a length-dependent small-fiber sensorimotor neuropathy first exhibited in the feet with ascending progression to the upper extremities. Autonomic neuropathy manifests as smooth muscle dysmotility (dysphagia, diarrhea, urinary retention), vascular dysregulation (orthostatic hypotension, erectile dysfunction), and anhidrosis. Soft tissue disease (carpal tunnel syndrome, tendonopathy, and spinal stenosis) commonly precedes nerve or heart manifestations of disease by 1–2 decades, particularly in ATTRwt amyloid patients who frequently report bicipital, patellar, or Achilles tendon rupture. Less common expressions of hATTR include vitreous opacities and leptomeningeal amyloid deposition from variant protein produced by the retinal epithelium and choroid plexus, respectively. ATTR amyloid involvement of the heart is clinically better tolerated than AL amyloid cardiomyopathy as reflected by the time from heart failure presentation to death in untreated cases of ATTR (median 42–48 months) versus AL (median 6 months) amyloidosis and the dramatically greater burden of disease by echocardiographic measures at symptomatic presentation.

Typical syndromes associated with other forms of AF disease include renal amyloidosis with mutant fibrinogen, lysozyme, or apolipoproteins; hepatic amyloidosis with apolipoprotein AI; and amyloidosis of cranial neuropathy with corneal lattice dystrophy pathognomonic of gelsolin amyloidosis. Patients with AF amyloidosis can present with clinical syndromes that mimic those of patients with AL disease. Rarely, AF carriers can develop AL disease or AF patients may have monoclonal gammopathy without AL. Thus, it is important to screen both for plasma cell disorders and for mutations in patients with amyloidosis. Although mass spectrometry often detects amino acid sequence variations, it is not designed to definitively identify specific protein variations; DNA sequencing is the diagnostic standard for AF mutations.

TREATMENT

ATTR Amyloidosis

Untreated, the survival period after onset of ATTR disease is 5–15 years. At present, three therapeutic strategies are used for ATTR amyloidosis: (1) orthotopic liver transplantation (OLT) to replace the factory of the mutated protein (only applicable to hATTR); (2) stabilization of circulating TTR tetramers, preventing TTR monomer release and amyloid fibril formation; and (3) TTR gene silencing (RNA interference or anti-sense oligonucleotide agents), suppressing hepatic TTR production to eliminate ATTR fibril formation. After 30 years of experience, OLT is largely limited to patients with hATTR amyloid and early peripheral neuropathy (V30M ATTR), as most patients with non-V30M TTR mutations suffer post-transplant progressive amyloid disease due to wild-type TTR from the allograft liver depositing on preexisting amyloid present in the heart and nerves. TTR tetramer stabilization successfully inhibits progressive ATTR amyloid nerve and heart disease as demonstrated by a phase 3 randomized controlled trial—the Diflunisal Trial (hATTR)—and the Transthyretin Amyloidosis Cardiomyopathy Clinical Trial (ATTR-ACT), respectively. Diflunisal, a repurposed generic nonsteroidal anti-inflammatory, and tafamidis, a proprietary thyroxine mimetic, bind TTR tetramers at the thyroxine binding site, minimizing release of the amyloidogenic TTR monomer and slowing progression of nerve and heart disease. Tafamidis, the first U.S. Food and Drug Administration-approved treatment for ATTR amyloid cardiomyopathy, extends survival and slows the decline in walking capacities and quality of life but does not appear to induce improvement in heart thickening or function. TTR gene silencers more reliably stop neurologic disease progression and, in 35–60% of treated patients with hATTR amyloid, improve sensory nerve deficits, a novel finding. Further, preliminary data suggest

TTR gene silencers may promote heart remodeling and improve systolic function.

The therapeutic future of ATTR amyloid patients is bright. Phase 3 randomized controlled clinical trials examining the safety and effectiveness of TTR gene silencers in patients with ATTR amyloid cardiomyopathy are underway, as are studies to determine the impact of second-generation TTR gene silencers on ATTR amyloid neuropathy and cardiomyopathy. TTR gene editing to prevent mRNA production or correct DNA mutations is the next frontier. Finally, as survival improves for patients with ATTR amyloid, therapies that cross the blood-brain barrier to address leptomeningeal (brain) and vitreous (eye) amyloid deposition arising from the choroid plexus and retinal epithelium, respectively, will be challenges to achieve.

$\text{A}\beta_2\text{M}$ AMYLOIDOSIS

$\text{A}\beta_2\text{M}$ amyloid is composed of β_2 -microglobulin, the invariant chain of class I human leukocyte antigens, and produces rheumatologic manifestations in patients undergoing long-term hemodialysis and, rarely, in patients with a hereditary form of disease. β_2 -Microglobulin is excreted by the kidney, and levels become elevated in end-stage renal disease. The molecular mass of $\beta_2\text{M}$ is 11.8 kDa—above the cutoff of some dialysis membranes. The incidence of this disease appears to be declining with the use of newer membranes in high-flow dialysis techniques. $\text{A}\beta_2\text{M}$ amyloidosis usually presents as carpal tunnel syndrome, persistent joint effusions, spondyloarthropathy, or cystic bone lesions. Carpal tunnel syndrome is often the first symptom. In the past, persistent joint effusions accompanied by mild discomfort were found in up to 50% of patients who had undergone dialysis for >12 years. Involvement is bilateral, and large joints (shoulders, knees, wrists, and hips) are most frequently affected. The synovial fluid is noninflammatory, and $\beta_2\text{M}$ amyloid can be found if the sediment is stained with Congo red. Although less common, visceral $\beta_2\text{M}$ amyloid deposits do occasionally occur in the gastrointestinal tract, heart, tendons, and subcutaneous tissues of the buttocks. There are no proven specific therapies for $\text{A}\beta_2\text{M}$ amyloidosis, but cessation of dialysis after renal allografting may lead to symptomatic improvement.

THERAPEUTIC FRONTIERS

To date, treatment strategies have focused on limiting formation of amyloidogenic proteins. Disruption of existing amyloid by targeting ubiquitous components of the tissue deposits offers theoretical means to improving major end-organ function; however, clinical trial validation remains elusive.

SUMMARY

A diagnosis of amyloidosis should be considered in patients with unexplained nephropathy, cardiomyopathy (particularly with diastolic dysfunction), neuropathy (either peripheral or autonomic), enteropathy, or the pathognomonic soft tissue findings of macroglossia or periorbital ecchymoses. Pathologic identification of amyloid fibrils can be made with Congo red staining of aspirated abdominal fat or of an involved-organ biopsy specimen. Accurate typing by a combination of immunologic, biochemical, and genetic testing is essential in selecting appropriate therapy (Fig. 112-1). Systemic amyloidosis should be considered a treatable condition, as anti-plasma cell chemotherapy is highly effective in AL disease and targeted therapies are being developed for AA and ATTR disease. The combination of precursor and end-organ amyloid therapeutics potentially provide not only disease control but also functional and quality of life improvements for patients with amyloidosis. Tertiary referral centers can provide specialized diagnostic techniques and access to clinical trials for patients with these rare diseases.

FURTHER READING

- ADAMS D et al: Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. *N Engl J Med* 379:11, 2018.
- BENSON MD et al: Inotersen treatment for patients with hereditary transthyretin amyloidosis. *N Engl J Med* 379:22, 2018.

- MAURER MS et al: Expert consensus recommendations for the suspicion and diagnosis of transthyretin cardiac amyloidosis. *Circ Heart Fail* 12:e006075, 2019.
- MAURER MS et al: Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *N Engl J Med* 379:1007, 2018.
- MERLINI G et al: Systemic immunoglobulin light chain amyloidosis. *Nat Rev Dis Primers* 4:38, 2018.
- SANCHORAWALA V: High dose melphalan and autologous peripheral blood stem cell transplantation in AL amyloidosis. *Hematol Oncol Clin North Am* 28:1131, 2014.
- SAROSIEK S, SANCHORAWALA V: Treatment options for relapsed/refractory systemic light-chain (AL) amyloidosis: Current perspectives. *J Blood Med* 10:373, 2019.

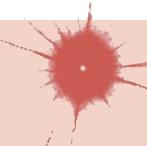
severe allergic reactions to BCs, or the manufacturing of pediatric units for young children and neonates.

BC constituents undergo centrifugation and filtration and are placed in contact with needles, plastic tubing and bags, as well anticoagulant molecules and various additive solutions. BCs are subjected to gas exchanges that are significantly different from aerobic breathing and are maintained at temperatures that are not physiologic, such as 22°C or 4°C. Any of these elements may contribute to so-called “storage lesions” that may occur at any time during BC processing and storage. Some of these lesions have proven to be reversible in the recipient after transfusion, while others may be irreversible. The clinical impacts of such lesions are under investigation. Storage lesions may also account for a number of adverse transfusion reactions, although there is currently no consensus on this issue.

Furthermore, plasma present in BCs contains donor antibodies (Abs). When directed toward antigens (Ags) present in the recipient, such as blood group or tissue (human leukocyte antigen [HLA]) Ags, such Abs may result in adverse events. RBCCs bring only a limited amount of donor plasma (10–30 mL), unlike PCs and obviously plasma. The use of platelet additive solution can replace two-thirds of plasma in PCs, while still leaving the equivalent of one plasma unit of 200 mL per transfused PC.

113

Transfusion Therapy and Biology



Pierre Tiberghien, Olivier Garraud,
Jacques Chiaroni

Transfusion encompasses the use of blood components (BCs) to prevent or treat anemia, hemorrhage, and bleeding disorders. Occasionally, BCs may be used to treat infection or relapse of malignant blood diseases after allogeneic hematopoietic transplantation. BCs comprise mainly red blood cell concentrates (RBCCs), platelet concentrates (PCs), and plasma for transfusion use (as opposed to plasma for fractionation into medicinal products such as albumin and immunoglobulin). Alongside transfusion safety, ensuring BC quality, assessing in vivo efficacy, and promoting evidence-based transfusion practices are critical aspects of transfusion medicine.

Blood collection and donor medicine do not fall within the scope of this chapter. Although the processes used are particularly safe, blood donations can cause adverse reactions, among which are fainting reactions and iron deficiency. These risks require preventive approaches and appropriate treatment when needed.

BLOOD COMPONENTS

BC collection and manufacturing processes are described in **Table 113-1**. Most common BCs are collected as whole blood or directly as components by apheresis. The vast majority of BCs are homologous. Autologous BCs, sometimes collected ahead of planned surgery, are now exceptional as they present little to no evidence-based advantage over homologous BCs. Nevertheless, such donation may still be of benefit in the presence of a rare blood group phenotype.

All BCs comply with common quality and performance standards and guidelines. Quality assurance encompasses well-defined processing steps and stringent BC quality controls as defined by health authorities. Tracing of all manufacturing steps as well as hemovigilance-based reporting of adverse events and incidents associated with blood collection, BC processing, and transfusion are highly recommended.

With the obvious exception of granulocyte concentrates and mononuclear cells, the majority of BCs are now leukocyte-reduced, and universal prestorage leukocyte reduction has been recommended. These BCs contain <1–5.10⁶ donor leukocytes and are associated with reduced incidence febrile nonhemolytic transfusion reactions (FNHTRs), infections with intracellular pathogens such as cytomegalovirus (CMV), alloimmunization, and immunomodulation.

BCs may undergo additional processing steps. These may include irradiation to prevent graft-versus-host disease (GVHD) in immunosuppressed patients, pathogen reduction to further reduce the risk of transfusion-transmitted infections, plasma reduction in patients with

BLOOD GROUP ANTIGENS AND ANTIBODIES

Red blood cells, as well as other blood constituents such as platelets and neutrophils, express allogeneic determinants. Transfusion may therefore result in alloimmunization and the production of Abs directed against allogeneic determinants. These alloantibodies (alloAbs) comprise anti-red blood cell (RBC) Abs, anti-HLA, anti-human platelet Ag (HPA) Abs, and anti-human neutrophil Ag (HNA) Abs. Anti-RBC immunization may result in hemolysis, while anti-HLA or anti-HPA Abs may result in other transfusion complications such as fever and platelet transfusion refractoriness. Furthermore, anti-HLA and anti-HNA immunization in the donor may result in a severe lung disorder called transfusion-related acute lung injury (TRALI). The Abs against red cell Ags may be IgM or IgG immunoglobulin classes. Some IgG or IgM can activate complement, and some IgG, crossing the placental barrier, may induce hemolytic disease of the fetus and newborn.

Erythrocyte blood groups refer to antigenic molecules that are expressed on the surface of RBC and other cells, genetically transmitted, and recognized by specific Abs. The polymorphism of such molecules explains their immunizing potential in situations such as transfusion, pregnancy, and transplantation. Blood groups can also interact with the environment and with infectious pathogens, leading to individual susceptibilities. For example, malaria is less severe in type O than non-O patients. Conversely, group O is associated with increased susceptibility to *Helicobacter pylori*. Currently, ~380 different blood group Ags have been described, classified within ~43 different systems. Blood group Ags belong to two broad categories based on their biochemical nature: carbohydrate blood groups and protein blood groups. RBC Ags may be the target of autoantibodies (autoAbs) generating autoimmune hemolytic anemia. Some of them, mostly IgG, are active at 37°C, called “warm autoAbs,” and are most often directed against Rh Ags, while others, most often IgM, are active at 4°C, called “cold autoAbs,” and may be directed against ABO, I, i, P, and other Ags.

Carbohydrate blood groups are headed by the **ABO system** which comprises two main Ags, A and B, encoded by two alleles, which are the A and B alleles, respectively. In addition to these active alleles, there is an inactive allele: O. Depending on the genotype, four different phenotypes are produced (**Table 113-2**). Other carbohydrate systems (H, P1PK, Lewis, I, and GLOB) share many characteristics with the ABO system. The main common feature is biochemical. Indeed, given their carbohydrate nature, Ags of the ABO system are considered to be “secondary products” of genes. The A allele encodes the A enzyme, which binds the A-type sugar (GalNac) A to the H substrate (expressed by action of the H enzyme encoded by the H allele, which happens to be inactive in the Bombay phenotype); sugars are attached to protein substrates on the surface of the RBC and so forth.

TABLE 113-1 Blood Components: Collection and Manufacturing Processes

BLOOD COLLECTION	INITIAL PROCESSING	BLOOD COMPONENT	ADDITIONAL COMPONENT PROCESSING (OPTIONAL TO MANDATORY)	RATIONALE	VOLUME AND CONTENT	STORAGE CONDITIONS AND DURATION
Whole blood	Separation into RBCCs and platelet-rich plasma (PRP) by slow centrifugation, followed by high-speed centrifugation of the PRP to yield one unit of platelets (most often subsequently pooled) and one unit of plasma. <i>Or</i> Separation into a PRBC, a plasma, and a "buffy coat" containing leukocytes and platelets by high-speed centrifugation, followed by pooling and slow-speed centrifugation of the buffy coat to produce a pooled platelet unit. Alternatively, the buffy coat may undergo high-speed centrifugation to produce a granulocyte unit that will be subsequently pooled.	RBCC from whole blood or from apheresis Platelets from whole blood (individual units or pools of 4–6 units of ABO identical units) or from apheresis	Deleukocytation to $<1\text{--}5.10^6$ leukocytes per unit; initial whole blood filtration or RBC elective filtration (highly recommended, mandatory in several international jurisdictions) Irradiation: X-ray or gamma, $\sim 25\text{--}35$ Gy; most often units no older than 28 days after collection Plasma reduction Pediatric preparation Cryopreservation (glycerol) Suspension in a platelet additive solution (PAS) Deleukocytation ($<1\text{--}5.10^6$ leukocytes per unit); initial whole blood filtration or platelet elective filtration (highly recommended, mandatory in several international jurisdictions) Pathogen reduction: Most often nucleic acid cross-linker and/or UV illumination	Reduction of posttransfusion fever and chills Reduction of intracellular pathogens (including CMV infections) Reduction of alloimmunization GVHD prevention in immunosuppressed patients or intrafamilial transfusions Prevention of allergic reactions in patients with prior severe reactions Adjustment to low-weight recipients Most often to ensure availability of RBCCs with a rare blood group for immunized "public-negative" recipients or recipients with complex alloimmunizations ^a Reduction of posttransfusion fever and chills Plasma orientation toward fractionation Reduction of posttransfusion fever and chills Reduction of intracellular pathogens (including CMV infections) Reduction of alloimmunization Reduction of transfusion-transmitted infections Prevention of GVHD	250–300 mL (including additive solution, no more than 40–50 mL of plasma) Hemoglobin: 22–40 g/dL Hematocrit: 50–70% Hemolysis $\leq 0.8\%$ at issuing Lesser volume, 10% reduction in RBC content Adjusted content Same Hb content Hematocrit: 40–80% Glycerol ≤ 1 g From 100 to 700 mL $\geq 2.10^{11}$ platelets Ph ≥ 6.4	4 \pm 2°C Duration depends on the additive solution: 25–42 days; some solutions aim to extend shelf life to 56 days After irradiation: 24 h After plasma reduction: 24 h to 10 days depending on reduction methodology N2 or -80°C electric freeze drying N2: unlimited; -80°C : 30 years 7 days after thawing in suitable additive solutions, 24 h if no additive solution At 20–24°C and under permanent motion: 3–7 days <i>Or</i> At 4°C without motion: up to 14–21 days (experimental) If irradiated: <24 h
Apheresis	Various apheresis devices allow for the collection of BCs either as individual BCs such as plasma or platelets (possibly double, such as double RBCC) or combined BCs, such as platelets and plasma, or RBCC, platelets, and plasma.		Volume reduction Irradiation: X-ray or gamma, $\sim 25\text{--}35$ Gy; in general, on bags no older than 3 days after collection Pediatric	Prevention of allergic reactions in patients with prior severe reactions Prevention of GVHD Volume and content adjustment		

(Continued)

TABLE 113-1 Blood Components: Collection and Manufacturing Processes (Continued)

BLOOD COLLECTION	INITIAL PROCESSING	BLOOD COMPONENT	ADDITIONAL COMPONENT PROCESSING (OPTIONAL TO MANDATORY)	RATIONALE	VOLUME AND CONTENT	STORAGE CONDITIONS AND DURATION
		Plasma from whole blood or from apheresis	Cryopreservation (DMSO) Cryopreservation at -18°C (most often) Deleukocytation (<1–5.10 ⁶ leukocytes per product); Initial whole blood filtration and/or plasma elective filtration Pathogen reduction: Nucleic acid cross-linker and/or UV illumination or solvent detergent treatment (most often on pooled products) Lyophilization	To ensure continuous availability in remote locations To ensure availability of platelets with rare HPA groups Shelf life extension Reduction of posttransfusion fever and chills Reduction of intracellular pathogens (including CMV) Reduction of alloimmunization Reduction of transfusion-transmitted infections	200–300 mL Coagulation factors, including fibrinogen (≥2 g/L), factor VIII (≥0.5 IU/mL), protein C and S, antithrombin	6 h after thawing (depending on cryopreservation procedure, may be resuspended in plasma) 1–2 years if cryopreserved Up to 28 days if kept unfrozen
		Granulocyte concentrates from whole blood (pools of up to x ABO identical units) or from apheresis ^b	Irradiation (mandatory)	To facilitate transportation and storage, as well as immediate availability, in remote locations Prevention of GVHD	≤650 mL ≤2.10 ¹⁰ granulocytes	Room temperature ≤24 h after the end of collection
		Whole blood	Deleukocytation with a platelet-sparing device	Reduction of posttransfusion fever and chills Reduction of intracellular pathogens (including CMV) Reduction of alloimmunization	~520 mL (including additive solution)	At 2–4°C Up to 25 days
		Peripheral blood mononuclear cells (apheresis)	May undergo cryopreservation (N2)	Increased practicability Repeated administration	Number of cells adjusted for a predetermined number of T lymphocytes 10 ⁵ –10 ⁷ CD3+ cells/recipient kg	N2: unlimited Never frozen or thawed: <6 h
		Cryoprecipitate (collected after thawing and centrifugation of plasma)	Resuspension in plasma (10–15 mL) and cryopreservation	N/A	Cold-insoluble plasma proteins (fibrinogen, factor VIII, von Willebrand factor)	12 months After thawing, may be stored at 20–24°C for up to 6 h

^aAntigen frequency below 1 to 4% (1/1000) of the population and contraindication for using regular blood units, depending on country-specific regulations. ^bGranulocyte collection by apheresis requires donor pre-administration of steroids and/or hematopoietic growth factor and exposure to heparin and HES during the apheresis procedure.

Abbreviations: BC, blood component; CMV, cytomegalovirus; DMSO, dimethyl sulfoxide; GVHD, graft-versus-host disease; Hb, hemoglobin; HPA, human platelet antigen; N2, nitrogen gas; N/A, not applicable; RBC, red blood cell; RBCC, red blood cell concentrate; UV, ultraviolet.

TABLE 113-2 ABO Blood Groups and Antibodies: Transfusion Compatibility

GENOTYPE(S)	ENZYME(S)/IMMUNODOMINANT SUGAR(S)	PHENOTYPE	NATURAL ANTIBODIES	TRANSFUSION COMPATIBILITY REQUIREMENTS		
				RBCC	PC ^a	PLASMA
A/A or A/O	"A" transferase/N-acetylgalactosamine (GalNAc)	A	Anti-B	A or O	A, O ^b , B ^b , or AB ^b	A or A,B
B/B or B/O	"B" transferase/galactose (Gal)	B	Anti-A	B or O	B, O, A ^b , or AB ^b	B or A,B
A/B	"A" transferase and "B" transferase GalNAc and Gal	A,B	None	A,B or A or B or O	A,B, O ^b , or A ^b or B ^b	A,B
O/O	Inactive Unconverted H antigen	O	Anti-A and Anti-B	O	O, A, B, or A,B	A or B or A,B or O

^aOrder of priority. ^bWithout high-titer anti-A and/or anti-B antibody.

Abbreviations: PC, platelet concentrate; RBCC, red blood cell concentrate.

Carbohydrate Ags are ubiquitously distributed in the body. The ABO Ags, expressed on endothelial cells, are genuine "tissue" groups and may be involved in graft rejection. These Ags are not specific to humans but are shared by many species including viruses and bacteria. The presence of A and B Ags in the environment and, in particular, on the bacteria of the microbiota explains the synthesis of so-called "natural" or "regular" Abs, aside from any transfusion or pregnancy. Such Abs have a major hemolytic capacity as they bind complement and activate its cascade up to the membrane attack complex. This imposes donor-recipient stringent compatibility rules for RBCCs and

whole blood transfusion and, albeit less stringently, for plasma and PC transfusion.

Protein blood groups are headed by the **Rh system** (formerly termed "Rhesus" or "Rh") for RBCs (**Table 113-3**). As these Ags are specific to humans, the occurrence of immunization can only occur upon allogeneic stimulation. The elicited Abs are called "immune" and "irregular" because their appearance following immunization is inconstant. These Abs directed against Ags of RBC groups other than ABO must be detected before RBCC cell transfusion or transplantation and during pregnancy. Of the 43 RBC group systems described, five

TABLE 113-3 Red Blood Cell (RBC) Group Systems and Antibodies: Clinical Significance and Transfusion Recommendations

ISBT NO./SYSTEM	SYMBOL/GENE(S)	ANTIGENS (NO.)	MAIN ANTIBODIES (ANTI-)	HEMOLYSIS CHARACTERISTICS		RBCC TRANSFUSION RECOMMENDATIONS
				TRANSFUSION	HDFN	
1/ABO	ABO/ABO	4	A, B	None to severe; immediate and/or delayed	None to moderate (rarely severe)	Ab-negative RBCC
2/MNS	MNS/GYPA, GYPB, (GYPE)	49	M	None (except in extremely rare cases if active at 37°C)	None (except in extremely rare cases if active at 37°C)	Compatible RBCC (negative DAT at 37°C)
			N	None (may be clinically significant in the case of the rare N-S-s-U- phenotype)	None	Ag-negative red cells in the case of sickle cell disease
			S, s U	None to moderate (rare) Mild to severe	None to severe (rare) Mild to severe (one reported case requiring an intrauterine transfusion)	Compatible RBCC (negative IAT at 37°C) Ag-negative RBCC in the case of N-S-s-U- phenotype Ag-negative RBCC Ag-negative RBCC
3/P1PK	P1PK/A4GALT	3	P1 P1, Pk, P (Tj ^a)	None to moderate; delayed (rare) None to severe	None None to severe	Compatible RBCC (negative DAT at 37°C) Ag-negative RBCC
4/Rh	RH/RHD, RHCE	55	D, C, E, c, e	Mild to severe; immediate or delayed	Mild to severe	Ag-negative RBCC
6/Kell	KEL/KEL	36	K	Mild to severe; delayed	Mild to severe (rare)	Ag-negative RBCC
7/Lewis	LE/FUT3	6	Le ^a , Le ^b	None (rare cases of hemolytic reactions)	None	Compatible RBCC (negative DIAT at 37°C)
8/Duffy	FY/ACKR1	5	Fy ^a , Fy ^b Fy3, Fy5	Mild to severe (rare); immediate/delayed Mild to moderate; immediate (rare)/delayed	Mild to severe (rare) Mild (rare) (no data for anti-Fy5)	Ag-negative RBCC Ag-negative RBCC
9/Kidd	JK/SLC14A1	3	Jk ^a , Jk ^b Jk3	None to severe; immediate or delayed None to severe; immediate or delayed	Mild to moderate (rare) None to mild	Ag-negative RBCC Ag-negative RBCC
18/H	H/FUT1	1	H (Bombay)	None to severe; immediate/delayed	Not none	Ag-negative RBCC
20/Globoside	GLOB/ B3GALNT1	2	P	None to severe	None to mild	Ag-negative RBCC

Abbreviations: Ab, antibody; Ag, antigen; DAT, direct antiglobulin test (Direct Coombs test); HDFN, hemolytic disease of the fetus and newborn; IAT, indirect antiglobulin test (Indirect Coombs test); ISBT, International Society of Blood Transfusion; RBCC, red blood cell concentrate.

(Rh, Kell, Duffy, Kidd, and MNS) are routinely investigated due to the clinical significance of Abs and their frequency. Testing for all five types ensures routine transfusion compatibility of 95%.

The Rh system comprises nearly 56 Ags, the most immunogenic of which is the RhD Ag (RH1). The Rh system has two *RH*D* and *RH*CE* genes located on chromosome 1. The *RH*D* gene codes for the RhD protein expressing the D Ag (RH1) present in 85%, 93%, and >99% of individuals of Caucasian, African, and Asian ancestry, respectively. The *RH*CE* gene codes for RhCE proteins expressing C (RH2) and/or c (RH4), and E (RH3) and/or e (RH4) Ags. The presence of the D Ag confers Rh “positivity,” while its absence confers Rh negativity. The *RH*D* and *RH*CE* genes determine eight main haplotypes (*DcE*, *DcE*, *Dce*, *DCE*, *dce*, *dCe*, *dCE*, and *dCE*) whose frequencies differ considerably among different geographical populations. The high diversity of the Rh Ags includes weak or partial expression. Identifying individuals (especially young females of childbearing potential and multitransfused patients) with a weak or partial RhD Ag is important to adequately select RhD-positive or -negative RBCs. Molecular biology is now routinely applied to resolve such situations.

The **Kell system** comprises 36 Ags, one of which is routinely determined: the K antigen (KEL1); 9% and 2% of individuals of Caucasian and African ancestry are K positive (KEL1), respectively, whereas 91% and 98%, respectively, are K negative (KEL-1). The immunogenicity of Kell is third behind the ABO and Rh systems. The Kell protein is linked to another blood group protein called Kx. The rare absence of this protein (controlled by a gene on X) is associated with a weak KEL Ag, acanthocytosis, shortened RBC survival, and a progressive form of muscular dystrophy that includes cardiac defects. This rare condition is called the McLeod phenotype.

The **Duffy system** (FY) comprises five Ags, two of which are routinely tested: the Fy^a Ag (FY1), coded by the *Fy^a* allele, and the Fy^b Ag (FY2), coded by the *Fy^b* allele. Depending on the combination of alleles, three common phenotypes are expected: Fy (a+b+), which has the two alleles *Fy^a* and *Fy^b*; Fy (a+b-), which has only the *Fy^a* allele in a double dose; and Fy (a-b+), which has only a double dose of the Fyb allele. A particular phenotype characterized by the absence of the Fy^a and Fy^b Ags, the Fy(a-b-) phenotype, is exclusive (with some exceptions) to individuals of African ancestry where it can reach frequencies of 70–100% depending on the population. It is linked to the presence of a double dose of a silent *FY*0* allele. This distribution may be related to the fact that the Fy Ags serve as receptors for *Plasmodium vivax* and therefore the Fy(a-b-) phenotype. However, these individuals may develop Abs against two high-frequency Ags (FY3 and FY5) after transfusion or pregnancy. They may also have low granulocyte counts that come to the attention of physicians, but the condition is not associated with any disease.

The **Kidd system** (JK) comprises three Ags, two of which are routinely tested: the Jk^a Ag (JK1), coded by the *Jk^a* allele, and the Jk^b Ag (JK2), coded by the *Jk^b* allele. Depending on the combinations of alleles, three common phenotypes are seen: Jk(a+b+) displaying the two alleles *Jk^a* and *Jk^b*, Jk(a+b-) displaying only the *Jk^a* allele in a double dose, and Jk(a-b+) displaying only a double dose of the *Jk^b* allele. A particular phenotype is characterized by the absence of the Jk^a and Jk^b Ags: the Jk(a-b-) phenotype found in Polynesian populations. It is linked to the presence of a double dose of a silent *JK*0* allele. These people may develop Abs against the high-frequency anti-JK3 Ag after transfusion or pregnancy.

The **MNS system** comprises 49 Ags, four of which are routinely tested. Two genes (*GYP4A*, *GYPB*) encode two pairs of so-called “anti-thetical” Ags. The M (MNS1) and N (MNS2) pair Ags encoded by the *M* and *N* alleles, respectively, are branched on the glycophorin A molecule. Their combination will determine whether or not they are present. M+ and N+ subjects have both alleles; an M+, N- subject is homozygous for the *M* allele; and an M-, N+ subject is homozygous for the *N* allele. The same holds true for the other pair of Ags, S (MNS3) and s (MNS4) expressed on glycophorin B. Therefore, an M+, N-, S-, s+ subject (in international nomenclature, this is written as MNS:1, -2, -3, 5) will be homozygous for the *M* and *s* alleles. A rare phenotype,

S-s-, found exclusively in individuals of African ancestry, can develop an Ab against the high-frequency U Ag (MNS:5) after transfusion or pregnancy.

RARE RBC PHENOTYPES

Some patients present with rare genotype/phenotype assortments and their RBCs display so-called private Ags or, conversely, lack public Ags (i.e., widely shared Ags) toward which the patient may develop an immune response when exposed to these Ags. Public-negative immunized individuals are virtually impossible to transfuse using conventional blood bank resources and require access to designated blood banks that have access to rare blood programs. Their primary responsibility is to identify and collect blood from donors exhibiting particular Ag displays on their RBCs or platelets that are uncommon in the given jurisdiction. Specific ethnic populations may be targeted, as some may display genotype specificities, such as the Bombay group in southwestern Indians. Several hemoglobinopathies, such as sickle cell disease, are more common in individuals of African ancestry. Such patients may display RBC phenotypes that are uncommon in countries in the Northern Hemisphere, resulting in difficulties adequately identifying donors to match the need, as a last resort, for highly valued cryopreserved BCs.

CLINICAL INDICATIONS AND EFFICACY ASSESSMENT OF BLOOD COMPONENTS

BCs are life-saving therapies but also scarce resources. Furthermore, transfusion may result in well-identified adverse reactions as well as more ill-defined adverse events, including inflammation and therapeutic inefficacy. As highlighted in so-called patient blood management programs, transfusion should be considered within a multidisciplinary approach that includes optimization of hematopoiesis, minimization of blood loss during surgical interventions, and optimization of tolerance to anemia. Clinical indications of BCs as well as means to assess therapeutic efficacy are detailed in **Table 113-4**.

ADVERSE REACTIONS TO BLOOD COMPONENTS

Adverse reactions to transfused BCs are most commonly non-life-threatening, although serious reactions can present with mild symptoms and signs. Transfused patients should be closely monitored for warning signs suggestive of adverse reactions, as described in **Table 113-5**. When an adverse reaction is suspected, the transfusion must be stopped while the recipient's clinical status is assessed and supportive care is initiated as needed. An average of 35 transfusion-associated fatalities with possible to definite imputability were reported yearly to the U.S. Food and Drug Administration (FDA) between 2014 and 2018 among ~14 million transfused BCs. Most frequent causes of death were transfusion-associated circulatory overload (TACO) (32%), followed by TRALI (26%), hemolysis (18%), and sepsis (14%).

Adverse reactions to BCs may result in immune and nonimmune mechanisms. Immune-mediated reactions are often due to recipient or donor alloimmunization and the presence of preformed recipient or donor Abs. Nonimmune causes of reactions are from the physical or chemical properties of BCs or from pathogens present in the BC.

IMMUNE-MEDIATED ADVERSE REACTIONS

Hemolytic Transfusion Adverse Reactions Immune-mediated acute hemolysis occurs when the recipient preformed Abs lyse transfused donor RBCs and may occur during or 24 h after transfusion. The anti-A or anti-B Abs are responsible for the majority of the most severe reactions, which can be fatal. However, alloAbs directed against other RBC Ags (i.e., Rh, Kell, and Duffy) are also responsible for severe hemolytic reactions. Such dramatic reactions are usually caused by a failure in product or patient identification, erroneous blood grouping, or unidentified anti-RBC alloimmunization in the recipient. Hemolysis, most often of lesser severity, may also occur upon transfusion of BCs containing incompatible plasma with a large amount of alloAbs directed against the recipient's RBCs. This may typically occur after

TABLE 113-4 Blood Components: Clinical Use

COMPONENT		THERAPEUTIC INDICATION	GOAL	DONOR/RECIPIENT COMPATIBILITY	DOSAGE	EFFICACY EVALUATION
Red blood cell concentrate (RBCC)	Transfusion	Anemia and/or tissue ischemia (treatment or prevention) Hb below a given threshold (to be considered in relation with clinical symptoms): <7 g/dL for patients hemodynamically stable, except for patients undergoing orthopedic surgery, cardiac surgery, or with preexisting cardiovascular disease (<8 g/dL) as well as for patients with acute coronary disease (<9–10 g/dL). Such thresholds do not apply to neonates and patients with severe thrombocytopenia and chronic transfusion-dependent anemia. Not recommended: nutritional anemia (iron, vitamin B ₁₂ , or folate deficiency)	Improve systemic and tissue oxygenation	ABO compatible (cellular) and ABO identical when achievable. RhD compatibility is required in young and childbearing females, and whenever possible if multitransfused RhC/c/E/e; Kell-compatible RBCCs are required in frequently transfused patients. Additional compatibility may be required depending on the clinical setting and screening results.	1 unit at a time (250–350 mL, including additive solution), repeated per clinical status and Hb level	Reduction of anemia-related symptoms, clinical improvement Increased Hb (+1 g/dL) and hematocrit (+3%)
	RBC exchange	Anemia/sickle cell crisis in hemoglobinopathies (sickle cell disease, thalassemia)	Replace altered RBCs with donor RBCs and compensate for hemolysis, prevention of sickle cell occlusive crisis		25–30 mL/kg	Sickle cell disease: reduced percentage of HbS
Platelet concentrates (PCs) (from pooled whole blood-derived platelets or single donor apheresis), maintained at room temperature (most often) or at 4°C		Thrombocytopenia-related bleeding disorders: treatment (cold or room temperature PC) or prevention (room temperature PC) Platelet level below a given threshold: ≤5000/μL in the absence of fever or infection, ≤10,000/μL to 20,000/μL if fever or infection; ≤50,000/μL if surgery, DIC, endoscopy, invasive procedures; ≤80,000/μL if neurosurgery or eye surgery Acute hypovolemic coagulopathy (see below) Not recommended: immune thrombocytopenia, thrombotic microangiopathy, heparin-induced thrombocytopenia	Correct impaired primary hemostasis, including vessel healing Cold stored platelets, despite lower in vivo survival, have maintained and possibly improved hemostatic capacity compared with room temperature stored platelets	ABO identical preferable; if not, ABO compatible (cellular) with low-titer anti-A/B Ab; RhD compatible preferred in premenopausal women HLA compatible (negative lymphocyte crossmatch) or HLA identical in case of refractoriness related to the presence of anti-HLA Ab HPA compatible in thrombocytopenic neonates to HPA immunized mother (fetal neonatal alloimmune thrombocytopenia)	0.5–0.7 × 10 ¹⁰ platelets/kg (apheresis or pooled whole blood-derived PCs)	Prevention and/or resolution of bleeding Corrected count increment ^a ≥10 × 10 ⁹ /L within 1 h and ≥7.5 × 10 ⁹ /L within 24 h after transfusion (not applicable to cold/cryopreserved platelets)
Plasma (thawed frozen, never frozen and maintained at 4°C or at room temperature, freeze-dried)	Transfusion	Coagulation factor-related bleeding disorders Acute hypovolemic coagulopathy (see below)	Correct impaired hemostasis by providing missing elements of coagulation or fibrinolysis cascade, as well as elements to heal injured vessel endothelium Provide Abs against relevant pathogens	ABO compatible (plasma)	10–15 mL/kg	Reduced bleeding disorder
	Plasma exchange (plasma or combined plasma and albumin)	Infectious disease treatment (convalescent plasma containing pathogen-specific Abs): Argentina hemorrhagic fever, viral respiratory infections (experimental) Pathogenic Ab removal and supplementation of lacking enzyme (e.g., thrombotic thrombocytopenic microangiopathy or Guillain-Barre syndrome) Pathogenic Ab removal (e.g., anti-HLA Ab prior to kidney transplantation)	Deplete pathogenic elements in the blood (auto-antibodies such as anti-ADAMTS-13 Ab in case of TTP, excess cholesterol, etc.); plasma may also bring anti-inflammatory and/or immunomodulatory factors such as immunoglobulin	ABO compatible (plasma)	Not determined 45–60 mL/kg	Infection resolution Improved disease-specific symptomatology (i.e., apyrexia and platelet recovery in case of TTP) Reduced antibody levels (e.g., anti-HLA antibodies prior to organ transplantation)

(Continued)

TABLE 113-4 Blood Components: Clinical Use (Continued)

COMPONENT	THERAPEUTIC INDICATION	GOAL	DONOR/RECIPIENT COMPATIBILITY	DOSAGE	EFFICACY EVALUATION
Whole blood	Acute hypovolemic coagulopathy requiring massive transfusion	Balanced provision of blood components maintained at 4°C and without an additive solution and related dilution	ABO-identical or group O with low-titer anti-A/B Ab	Repeated per clinical status	Normovolemia; bleeding resolution
Multicomponent (RBCC, PC, and plasma)	Acute hypovolemic coagulopathy requiring massive transfusion	Appropriate ratio is under investigation; a ratio of 1 RBCC/1 plasma/0.25 PC (platelet content of a whole blood) is currently favored	Standard RBCC, PC, and plasma compatibility	1 RBCC/1 plasma/0.25 PC ratio, repeated per clinical status	Normovolemia; bleeding resolution
Granulocyte concentrates (apheresis or a pool of whole blood-derived granulocytes)	Severe refractory bacterial or fungal infection in patients with neutropenia (<100/ μ L) or with dysfunctional granulocytes (CGD) (mainly soft tissues and lung). Neutropenia can be acquired (chemotherapy) or congenital. Usefulness of granulocyte transfusions is debated. Formal proof of efficacy is lacking.	Correct impaired granulocyte function in relation to granulocytopenia or granulocyte dysfunction	ABO compatible	1–2 $\times 10^{10}$, repeated per clinical status	Infection resolution (or stabilization until recovery from neutropenia)
Donor mononuclear cells	Relapse of malignant hemopathy after allogeneic hematopoietic cell transplantation	Graft-versus-leukemia effect (and graft enhancement effect)	N/A	10 ⁶ –10 ⁷ T lymphocytes/kg	Disease specific (remission)
Cryoprecipitate	Acute bleeding coagulopathy, type II (dysfunctional factor) or type III (absent factor) Von Willebrand disease, hemophilia A in the absence of factor VIII concentrates	Provision of fibrinogen, factor VIII, von Willebrand factor, and factor XIII	ABO compatibility is not required	10–15 mL/unit, pool of 4–5 units	Increased plasma fibrinogen (0.3–1 g/L)

^aCCI calculation:

$$\text{CCI} = \frac{\text{Postransfusion count } (/ \mu\text{L}) - \text{pretransfusion count } (/ \mu\text{L})}{\text{Number of platelets transfused} \times 10^{11}} \times \text{Body surface area } (\text{m}^2)$$

Abbreviations: Ab, antibody; CCI, corrected count increment; CGD, chronic granulomatous disease; DIC, disseminated intravascular coagulation; Hb, hemoglobin; HLA, human leukocyte antigen; N/A, not applicable; RBC, red blood cell; TTP, thrombotic thrombocytopenic purpura.

TABLE 113-5 Transfusion Adverse Reactions: Main Warning Signs

Fever ($\geq 38^\circ\text{C}$)	+1–2°C within 4 h +1–2°C within 15 min +/-: • Chills • Dyspnea • Hypotension • Digestive disorders • Disseminated intravascular coagulation • Hemoglobinuria $>2^\circ\text{C}$ or $\geq 39^\circ\text{C}$	FNHTR Anti-HLA immunization and cognate Ag in the blood product TRALI (with dyspnea at the forefront) Transfusion-transmitted bacterial infection Hemolysis
Hypotension (≥ 30 mmHg decrease in systolic blood pressure)		Hemolytic shock Anaphylactic shock Septic shock TRALI (with dyspnea at the forefront)
Dyspnea		TRALI (within 6 h of transfusion) TACO (within 6 h of transfusion) Severe allergy (immediate; within 4 h)
Hemoglobinuria		Intravascular hemolysis • Immunologic • Mechanical • Toxic • Thermic
Rash	<2/3 of the body within 2–3 h >2/3 of the body during or within 2–3 h >2/3 of the body within 5 min Associated with dyspnea and shock	Minor allergy Severe allergy Anaphylaxis
Icterus		Delayed hemolysis
New alloantibody		Alloimmunization
Rash, diarrhea, and fever occurring 2 days to 6 weeks after transfusion		GVHD
Gum bleeding, purpura 5–12 days after transfusion		Posttransfusion purpura
Cardiac, hepatic, and/or renal insufficiency in frequently transfused patients		Posttransfusion iron overload
Top-down investigation after a blood donor is subsequently found to be infected Bottom-up investigation after another recipient of a same blood donation is found to be infected Infectious symptoms within 6 months		Transfusion-transmitted infection

Abbreviations: Ag, antigen; FNHTR, febrile nonhemolytic transfusion reaction; GVHD, graft-versus-host disease; HLA, human leukocyte antigen; TACO, transfusion-associated circulatory overload; TRALI, transfusion-related acute lung injury.

transfusion of a PC containing ABO-incompatible plasma. Estimated frequencies of acute and chronic hemolytic adverse reactions are 1–10 and 5–40 per 10^5 transfused BCs, respectively.

Mechanisms of transfusion hemolytic reactions are described in **Figure 113-1**.

Prevention of hemolytic reactions relies on **pretransfusion testing** of potential recipients. Testing will include determination of the ABO RhD phenotype (and anti-ABO Abs) as well as additional typing for the other main Rh Ags (CcEe): K Ag of the Kell system and, more rarely, Duffy, Kidd and Ss Ags, depending on the clinical setting. These determinations are most often performed by serology. However, molecular typing is increasingly being used to predict RBC phenotype and facilitate the selection of a compatible component. Special care must be taken to verify the patient's identity and apply adequate tube labeling. A double ABO determination performed separately may be considered, especially in the absence of a systematic crossmatch.

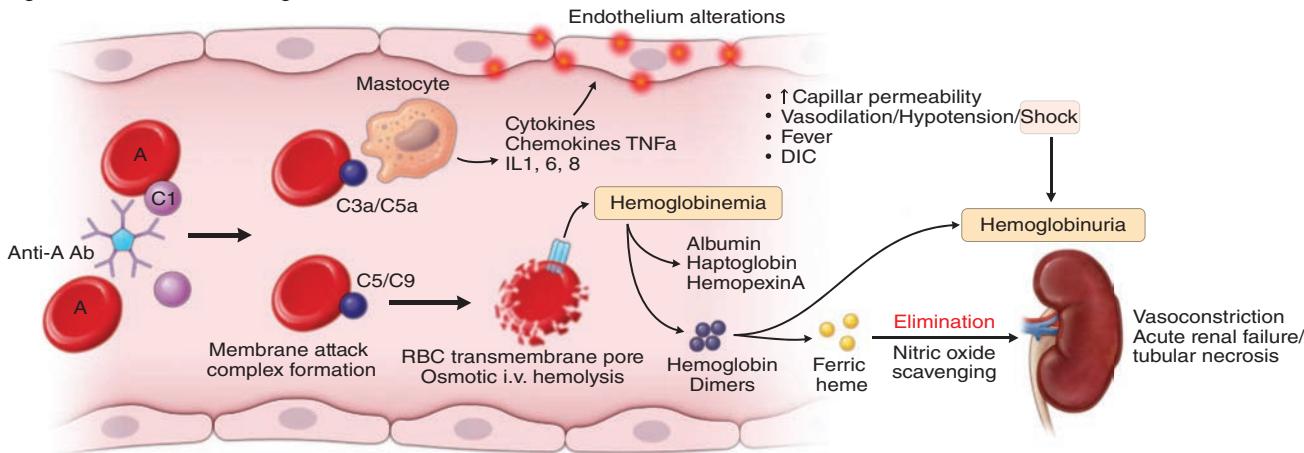
Testing will also include the screening and identification of alloAbs directed against RBC Ags other than ABO. This screen is performed by mixing patient serum with type O RBCs expressing Ags from most blood group systems and whose extended phenotype is known. The specificity of the alloAb is identified by correlating the presence or absence of Ag with the induced—or not—agglutination. Special attention should be paid to patients receiving monoclonal Ab treatment that

may bind to erythrocytes *in vivo* (such as anti-CD38 IgG treatment for multiple myeloma) and therefore interfere with alloAb screening. Such interference may be offset by sample dithiothreitol pretreatment.

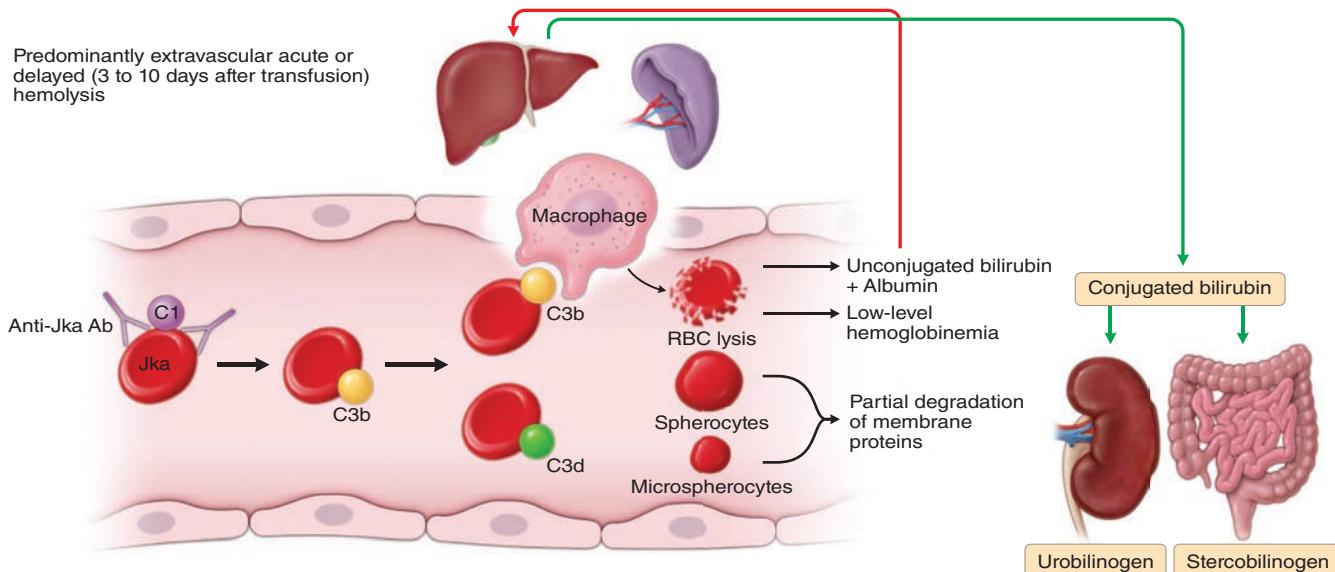
Crossmatching between the recipient plasma/serum and the sample of selected RBCs may be performed, especially when the recipient is alloimmunized against RBC or is frequently transfused, as well as in specific clinical settings such as sickle cell disease, even if the Ab screening is negative.

The selection of a compatible BC should take into account pretransfusion testing as well as the recipient's clinical status. In the case of D (Rh1)-negative patients, every attempt must be made to provide Rh-negative BC to prevent anti-D alloimmunization. In an emergency situation, D-positive RBCC can be safely transfused to a D-negative patient who lacks anti-D. However, an estimated 20–22% of RBCC recipients will become alloimmunized and produce anti-D Abs after transfusion with D-positive RBCs (this frequency is higher in healthy individuals). Such alloimmunization can occur after PC transfusion, although at a much lower frequency (~1%). Whenever possible, females with childbearing potential (to include prepubertal girls) should be transfused with D- and K (KEL1)-compatible RBCCs and D-compatible PCs to prevent alloimmunization and protect a future fetus/newborn from an alloimmune-mediated hemolytic disease. D-negative females with childbearing potential who are transfused

A Predominantly intravascular acute hemolysis occurring during or within 24 hours following transfusion



B Predominantly extravascular acute or delayed (3 to 10 days after transfusion) hemolysis



C Predominantly extravascular acute or delayed (3 to 10 days after transfusion) hemolysis

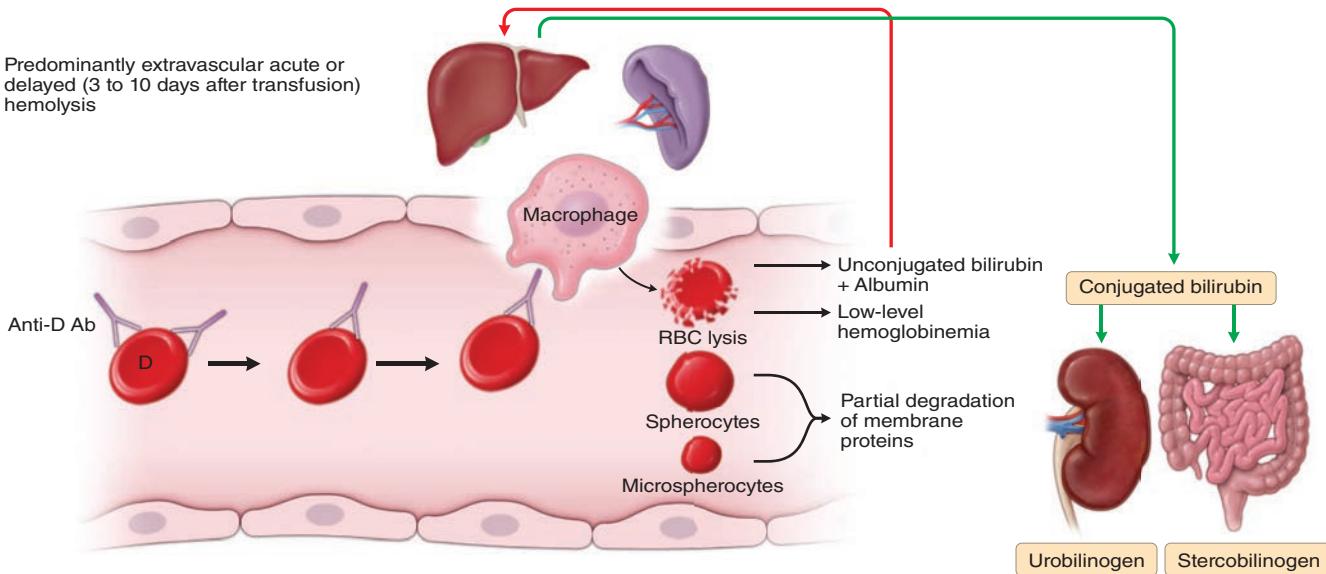


FIGURE 113-1 Mechanisms of transfusion hemolytic reactions. **A.** Acute responses will involve preexisting antibodies (Abs), naturally occurring anti-A/anti-B IgM or IgG directed against other RBC Ab and resulting from prior sensitization. Upon interaction with cognate antigen (Ag) on transfused red blood cells (RBCs), recipient alloantigenic Ab (alloAb), mostly natural anti-A/anti-B IgM, may fix and activate complement up to C5/C9. Formation of membrane attack complex (MAC) will create pores in transfused RBCs with resulting intravascular hemolysis, release of toxic moieties including free hemoglobin responsible for end-organ damage including renal failure, and tissue factors contributing to occurrence of disseminated intravascular coagulation (DIC). **B.** Alternatively, complement activation may be incomplete, as typically observed in a delayed hemolytic transfusion reaction involving neoformed alloantigenic IgG. In such cases, complement activation up to C3 results in C3b-mediated opsonization of RBCs, extravascular hemolysis, and clearance through immunophagocytosis. Anemia and jaundice will be the primary clinical manifestations. **C.** Lastly, alloAb may not fix complement while ensuring antibody-dependent cellular cytotoxicity (ADCC)-mediated phagocytosis of targeted RBC. (Adapted from SR Panch et al: Hemolytic transfusion reactions. *N Engl J Med* 381:150, 2019.)

with BCs containing Rh-positive RBCs should receive anti-D Ig to prevent allo sensitization.

Hemolysis, most often of lesser severity, may also occur after transfer of alloAbs directed against the recipient's RBC Ags. Such ABO "plasmatic" incompatibility, called "minor ABO incompatibility," will occur mainly with PC transfusions, where platelets are suspended in ~100–300 mL of plasma (depending on whether part of the plasma is substituted by additive solution). BCs containing plasma with high-titer anti-A/B Ab may induce a hemolytic reaction. When the transfusion of ABO-identical (vs ABO-compatible) PCs is feasible, PCs provided by donors with low-titer anti-A/B only should be preferred. "High-titer" PCs should be restricted to group O recipients. While there is no universal definition of high-titer Abs, a threshold titer of 1/64 (as assessed by hemagglutination) may be appropriate. It should be noted that the use of an additive solution in PCs substantially mitigates this risk. Lastly, ABO plasmatic incompatibility can lead to the formation of immune complexes with soluble A and/or B Ags and ensuing inflammation and platelet activation.

Acute hemolytic reactions may present with hypotension, tachypnea, tachycardia, fever (+1–2°C), chills, chest and back pain, hemoglobinuria, and hemoglobinemia. In the most severe cases, DIC, acute renal failure, shock, and death may occur.

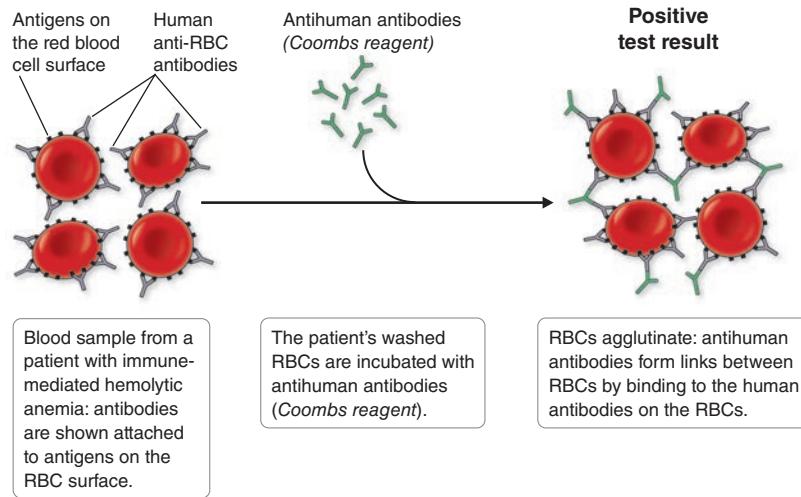
Delayed hemolytic reactions, with icterus and persisting or worsening anemia as the main clinical manifestations, result from an anamnestic response. Such reactions may occur in patients previously

sensitized to RBC Ags who have a negative alloAb screen at the time of transfusion due to low Ab levels. The alloAb is detectable 1–2 weeks after the transfusion.

Diagnosis of transfusion-associated hemolysis relies on persistent and/or worsening anemia, depleted plasma haptoglobin levels, hemoglobinuria and hemoglobinuria, as well as elevated plasma lactate dehydrogenase and unconjugated bilirubin. The direct antiglobulin test (DAT, or direct Coombs test) that detects immunoglobulin, and possibly complement (C3d), on the surface of the recipient's RBC will most often be positive (Fig. 113-2). Similarly, a positive indirect antiglobulin test (IAT, or indirect Coombs test) that detects anti-RBC alloAb in the serum will also be positive. An elution of the Ab on the surface of the RBC may allow for the identification of the culprit alloAb.

The management of an immune-mediated acute hemolytic transfusion reaction is mainly supportive. Prompt interruption of the transfusion, biological workup, and a thorough clerical check to prevent a possible second misidentified transfusion are crucial initial steps. Vigorous hydration with isotonic saline and diuretics to maintain urine output is recommended. Although often self-limiting, acute hemolysis may also require forced alkaline diuresis, correction of electrolyte abnormalities, and pressor support as needed. In patients with DIC and severe bleeding, PC, plasma, and cryoprecipitate or fibrinogen may be required. When transfusion of incompatible RBCCs is unavoidable, prophylaxis with steroids (100 mg of hydrocortisone) just before the transfusion and repeated 24 h later and polyvalent immunoglobulin

Direct Coombs test/direct antiglobulin test



Indirect Coombs test/indirect antiglobulin test

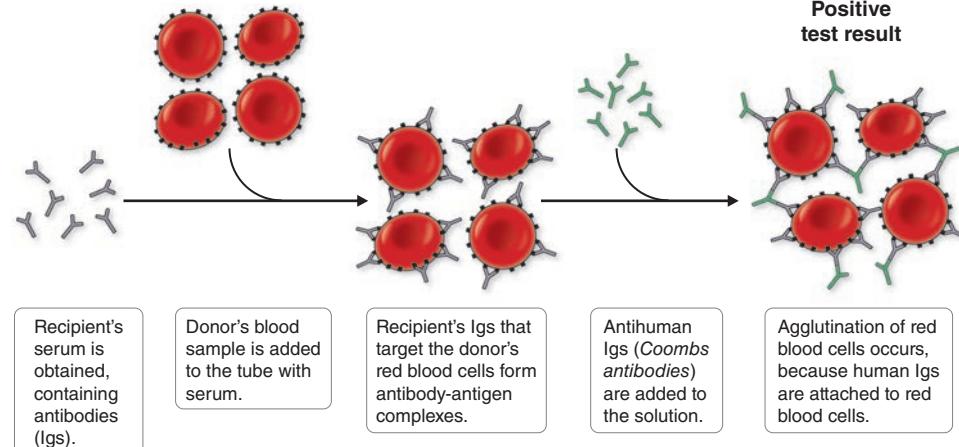


FIGURE 113-2 Direct and indirect Coombs test. The direct Coombs (antiglobulin) test detects the presence of antibodies (or complement) on the surface of erythrocytes. The indirect Coombs (antiglobulin) test detects antibodies in the serum that may bind to donor erythrocytes. Igs, immunoglobulins; RBC, red blood cell. (Adapted from http://upload.wikimedia.org/wikipedia/commons/1/1c/coombs_test_schematic.png.)

(1.2–2.0 g/kg per day over 2–3 days, initiated just before the transfusion) have been successfully used to prevent or minimize acute and delayed hemolysis.

Immune-mediated hemolysis may also occur after allogeneic hematopoietic transplantation (most often involving a peripheral blood stem cell graft) or, more seldomly, solid organ transplantation. Minor ABO incompatibility, with subsequent red cell destruction in the recipient, is the most common cause of clinically significant hemolysis in such cases. Viable donor B lymphocytes, called “passenger lymphocytes,” transferred passively with the graft, may produce alloAbs (including anti-D or anti-A1 in an A2 donor) that target recipient red cells. Such hemolysis has been reported to develop 5–14 days after transplantation. Reduced-intensity conditioning regimens and cyclosporine as prophylaxis against GVHD or rejection have been associated with increased risk. Transfusing RBCs compatible with the graft donor and the use of GVHD prophylaxis able to target B cells (e.g., methotrexate) have significantly reduced the incidence of passenger lymphocyte syndrome. Allogeneic hematopoietic transplantation may also result in acute hemolysis due to incompatible donor-derived red cell (and precursor) destruction by the recipient alloAbs (i.e., major ABO incompatibility). Prolonged pure red cell aplasia may occur in such a situation. Graft deserythrocytation will reduce the risk of early acute hemolysis.

Polyvalent immunoglobulin may contain high titers of anti-A (mostly) and/or anti-B Abs and induce acute hemolysis, most often of limited severity. Such hemolysis is particularly described in group A or A,B children receiving high-dose immunoglobulin, notably for Kawasaki's disease, as well as in adults treated for thrombotic thrombocytopenic purpura. A similar mechanism may lead to hemolysis after anti-D immunoglobulin treatment for immune thrombocytopenia in RhD-positive patients.

Nonimmune mechanisms of transfusion-associated hemolysis include thermal (overheated or cold BCs), osmotic (concurrent hypo-osmotic perfusion), and mechanical (pressure related to high-flow transfusion filtering during cell saver processing) mechanisms.

Autoimmune and drug-induced hemolytic anemias may be exacerbated by transfusion and can therefore mimic hemolytic transfusion reactions. Transfusion of RBCs with enzymatic defects may mimic immune-mediated hemolysis as well. Notably, severe hemolytic reactions in patients receiving long-term transfusions for hemoglobinopathies (mainly sickle cell disease) can precipitate bystander hemolysis, in addition to clearing transfused red cells. The mechanisms of this hyperhemolytic transfusion reaction may be a mediated RBC hemolysis-related systemic inflammatory response and resulting lysis of red cell precursors by macrophages. This process may be immediate or delayed, with hemoglobin levels falling below the pretransfusion values, often to life-threatening levels. Further RBCC transfusion typically exacerbates ongoing hemolysis, with the exogenous (transfused) allogeneic Ags probably triggering further nonspecific hemolysis.

Febrile Nonhemolytic Transfusion Reaction The most frequent reaction associated with the transfusion of cellular BCs is FNHTR. This reaction is characterized by chills and rigors and a $\geq 1^\circ\text{C}$ rise in body temperature and is caused by proinflammatory cytokines in the BC or by recipient Abs directed against donor cell Ags present in the BC. FNHTR is diagnosed when other causes of fever, notably infection and hemolysis, have been excluded in the transfused patient. Leukocyte reduction, especially prestorage, can prevent the occurrence of FNHTR. Moreover, the use of additive solutions decreases FNHTR frequency associated with PC transfusion. Premedication with antipyretics has generally proven ineffective at decreasing the rate of such reactions and may mask relevant clinical symptoms.

Allergic Reactions Most allergic transfusion reactions are mild and include rash, pruritus, urticaria, and localized edema. More rarely, allergic reactions may be severe to life-threatening with an anaphylactic reaction that can involve bronchospasm, respiratory distress, hypotension, nausea, vomiting, and shock. Frequencies of mild and severe allergic reactions are ~ 100 and ~ 5 per 105 BCs, respectively.

Allergic reactions are related to plasma proteins found in transfused components. Mild reactions may be treated by temporarily stopping the transfusion and administering antihistamine drugs. Patients with a history of allergic transfusion reaction may be premedicated with an antihistamine, although there is no consensus on this issue. Cellular components can be washed to remove residual plasma for extremely sensitized patients. Most of the allergic presentation may not depend on preformed Abs and may be attributable to soluble mediators triggering histamine and serotonin release from platelets and leukocytes. An anaphylactic reaction may occur after the transfusion of only a few milliliters of the BC. Treatment includes stopping the transfusion, maintaining vascular access, and administering adrenaline (0.3–0.5 mg subcutaneously). Additional treatment with steroids, antihistamine drugs, and bronchodilators may also be required.

Patients who are IgA deficient ($<1\%$ of the population) may be sensitized to this immunoglobulin isotype and may be at risk of anaphylactic reactions associated with plasma transfusion. As a precaution, individuals with severe IgA deficiency should therefore receive, where available, IgA-deficient plasma and washed cellular BCs. Patients who have anaphylactic or repeated allergic reactions to BCs should be tested for IgA deficiency. It should be noted that the importance, or even the reality, of such a transfusion-related allergic risk is currently debated.

Graft-Versus-Host Disease GVHD is an extremely rare adverse reaction caused by transfusion, although it is a frequent complication of allogeneic hematopoietic transplantation. Transfusion-related GVHD is mediated by engrafted donor T lymphocytes in a recipient unable to reject such allogenic lymphocytes (as in severely immunosuppressed patients or patients homozygous for an HLA haplotype shared with the donor). Such donor T lymphocytes interact with host HLA Ags and mount an immune response, which is manifested clinically by the development, 5–10 days after transfusion, of cytopenia, fever, a characteristic skin rash, diarrhea, and liver function abnormalities. Transfusion-associated GVHD is highly resistant to treatment with immunosuppressive therapies as well as ablative therapy followed by allogeneic bone marrow transplantation and is fatal in $>90\%$ of cases. Prevention in at-risk patients relies on the irradiation of cellular BCs (minimum of 25 Gy) or treating BCs with pathogen reduction technology that will deplete all living cells in the component. At-risk patients include patients with inherited immune deficiency, patients undergoing autologous or allogeneic hematopoietic transplantation, patients treated with immunosuppressive drugs such as purine or pyrimidine analogues, anti-CD52 Ab or antithymocyte globulin, fetuses receiving intrauterine transfusions, and recipients of BCs provided by a blood relative. Because granulocyte concentrates contain a large number of lymphocytes, they should always be irradiated.

Transfusion-Related Acute Lung Injury TRALI is characterized by the occurrence or worsening of hypoxia and noncardiogenic pulmonary edema with bilateral interstitial infiltrates on chest x-ray during or within 6 h after transfusion, although delayed cases may occur up to 72 h later. Frequency of TRALI is BC dependent and ranges, on average, from 0.5 to 10 per 10^5 BCs. TRALI may be difficult to distinguish from other causes of hypoxia, such as circulatory overload, and is among the most common causes of transfusion-related fatalities. Treatment is supportive only. TRALI usually results from the transfusion of donor plasma that contains high-titer anti-HLA class II Abs that bind recipient cognate Ag. Anti-HLA class I and anti-human neutrophil antigen (HNA) Abs may also be involved. TRALI mediated by cytokines and chemokines in the absence of an HLA-mediated interaction may occur also. Leukocytes, especially when primed by either a bacterial moiety such as lipopolysaccharide or a cytokine/chemokine, aggregate in the pulmonary vasculature and release inflammatory mediators. The transfusion of plasma and PCs from male donors and nulliparous or parous female donors without anti-HLA Abs has significantly reduced the risk of TRALI where implemented. Recipient factors associated with an increased risk of TRALI include smoking, chronic alcohol use, shock, liver surgery (transplantation), cancer surgery, mechanical ventilation, and positive fluid balance.

Posttransfusion Purpura This rare reaction ($\sim 1/10^5$ BCs) is defined as a thrombocytopenia-related bleeding disorder developing 5–12 days after PC (and more rarely RBCC) transfusion, predominantly in women. Platelet-specific alloAbs are found in the recipient, most frequently anti-HPA-1a in HPA-1a-negative alloimmunized individuals. The delayed thrombocytopenia is due to a secondary increased production of alloAbs. The mechanisms for the destruction of the patient's own platelets remain unclear. Management is mostly supportive but may require polyvalent immunoglobulin, steroids, or plasma exchange. Additional platelet transfusions may worsen the thrombocytopenia or be associated with poor increments. Prevention of recurrence includes use of washed BCs or BCs from HPA-compatible donors.

Alloimmunization/Platelet Refractoriness A recipient may become alloimmunized to a number of Ags on cellular blood elements and plasma proteins. AlloAbs to RBC Ags are detected during pretransfusion testing, and their presence may delay finding Ag-negative crossmatch-compatible products for transfusion. Women of childbearing age who are sensitized to RBC Ags (i.e., D, c, E, Kell, or Duffy) are at risk of bearing a fetus with hemolytic disease of the fetus or newborn. Ag matching is the only pretransfusion selection test to prevent RBC alloimmunization, which is found to occur with a frequency of $\sim 100/10^5$ RBCC transfusions. Alloimmunization to Ags on leukocytes and platelets, most often anti-HLA Abs, can result in refractoriness to PC transfusions (as defined by a low increase in platelet count after transfusion). Once alloimmunization has developed, HLA-compatible (crossmatched) PCs should be preferred if available. If not, repeated PCs at shortened intervals may be considered. Use of leukocyte-reduced cellular BCs will reduce the incidence of immunization. Transfusion refractoriness may also result from an anti-HPA alloimmunization, although less commonly. Recipient factors associated with platelet refractoriness include fever, splenomegaly, bleeding, DIC, and medications such as amphotericin B. Notably, cold-stored (and cryopreserved) PCs have been found to have preserved hemostatic function in acutely bleeding patients despite poor platelet increments.

Immunomodulation Transfusion of allogeneic blood may be associated with immunosuppression, as evidenced early on by the beneficial effect of pretransplant transfusion on kidney graft survival. The intensity of such an effect is debated and, if present, is most probably attenuated by the use of leukoreduced BCs. Transfusion-related immunomodulation is indeed thought to be mainly mediated by donor leukocytes, whether transfused to the recipient or undergoing apoptosis during storage. However, leukoreduced RBCCs or PCs still release immunomodulatory mediators during storage. These mediators, along with the transfused RBCs or platelets, may exert various, possibly opposing, immune effects *in vivo*, including immunosuppression and inflammation.

■ NONIMMUNOLOGIC TRANSFUSION ADVERSE REACTIONS

Fluid Overload TACO is a common and underrecognized transfusion adverse reaction. Estimated frequencies vary from ~ 10 to 1000 per 10^5 BCs. TACO is now the main cause of death from transfusion since the TRALI risk has been mitigated. Risk factors include older age, renal failure, preexisting fluid overload, cardiac dysfunction, administration of a large volume of BCs, and an excessive rate of transfusion in relation to the patient's hemodynamic tolerance. TACO results in dyspnea, hypoxia, bilateral and predominantly alveolar infiltrates on chest x-ray, frequent systolic hypertension, and elevated brain natriuretic peptide. Fever may also exist. Prevention involves identifying at-risk patients, close monitoring, a slow transfusion rate (1 RBCC over 3–4 h), and use of diuretics in hemodynamically stable patients with a history of TACO. Treatment requires stopping the transfusion and administering oxygen and diuretics.

Massive Transfusion-Associated Reactions/Electrolyte and Cold Toxicity Reactions Reactions related to massive transfusion,

i.e., transfusion of 50% of the patient's total blood volume over 3 h or >5 –10 units of RBCCs (plus associated BCs), include citrate toxicity, hypothermia, hyperkalemia, and dilutional coagulopathy. Citrate, which is commonly used to anticoagulate BCs, chelates calcium. Hypocalcemia, manifested by circumoral paresthesia, and changes in cardiac function may result from multiple rapid transfusions. Although citrate is quickly metabolized to bicarbonate, calcium infusion (through a separate line) may be required. Rapid transfusion of BCs still at 4°C can result in hypothermia and cardiac dysrhythmias. Use of an inline warmer will prevent this complication. RBC leakage during storage, longer storage, and irradiation increase the concentration of potassium in the unit. Neonates and patients with renal failure or other comorbidities (e.g., hyperglycemia or hypocalcemia) are at risk of hyperkalemia and resulting acute cardiac toxicity. Treatment includes insulin, glucose, calcium gluconate, and furosemide, and prevention includes the use of washed or plasma-reduced RBCCs or a storage age of <7 –10 days and the avoidance of RBCCs stored for >24 h after irradiation.

Iron Overload Each unit of RBCs contains 200–250 mg of iron. In frequently transfused recipients, iron accumulation that is left untreated will affect endocrine, hepatic, and cardiac function. Death may occur from cardiac failure or arrhythmia. Iron overload can be assessed by means of serum ferritin measurements, magnetic resonance imaging, and liver biopsy. Prevention and treatment of this frequently underreported transfusion adverse event rely on careful monitoring and iron chelation.

Hypotensive Reactions Acute hypotensive transfusion reactions are defined as an abrupt drop in blood pressure of >30 mmHg early after the start of transfusion and resolving quickly once the transfusion is stopped, without further intervention. Respiratory, gastrointestinal, or mild allergic reactions may also be present. Estimated frequency is 1 – $10/10^5$ BCs. These reactions may result from the generation of vasoactive kinins in the BCs and are more likely to occur in hypertensive patients taking angiotensin-converting enzyme (ACE) inhibitors who are therefore less able to metabolize bradykinin. Upon resolution, the same blood product should not be restarted. Switching from an ACE inhibitor to an alternative drug should be considered for patients requiring further transfusions.

Adverse Transfusion Reactions of Uncertain Imputability Necrotizing enterocolitis, which is common in preterm and very-low-birth-weight neonates, has been infrequently described with close temporal association with RBC transfusion. However, the causality of any association remains to be further ascertained, as does the efficacy of withholding feeds during transfusion to prevent such a complication. Posterior reversible encephalopathy syndrome is a rare syndrome characterized by acute reversible neurologic symptoms related to subcortical vasogenic brain edema. It has been described within 10 days after RBCC transfusion, mainly in women with severe (and long-standing) anemia. The prognosis is most often favorable, although irreversible neurologic disturbance has been described. Prevention may include avoiding rapid correction of chronic severe anemia. Again, causality remains to be established.

■ INFECTIOUS ADVERSE REACTIONS

Donor screening involves the selection of healthy donors without high-risk lifestyles, medical conditions, or exposure to transmissible pathogens. Tests are performed on donated blood to detect the presence of infectious agents by testing for relevant Abs or by directly detecting infectious agents most often by nucleic acid amplification testing. The increasing sensitivity of testing methods has progressively narrowed the “window” period early on after infection during which a low-titer undetectable virus may be present in the blood and result in a transfusion-transmitted infection.

Transfusion-transmitted bacterial infection remains a significant concern, notably with PCs stored at room temperature, which allows for bacterial proliferation and results in an increased risk during storage. However, some gram-negative bacteria such as *Yersinia* can grow

TABLE 113-6 Infectious Transfusion Adverse Events

PATHOGEN		DONATION PREVALENCE (/10 ⁴ BLOOD DONATIONS)	PREVENTION MEASURES (IN ADDITION TO DONOR DEFERRAL)	INFECTION PREVALENCE IN RECIPIENTS (/10 ⁶ BLOOD PRODUCTS TRANSFUSED)
Bacteria	Pyogenic bacteria	PC: 10–20	Asepsis, diversion of the initial 10–30 mL of blood, bacterial detection, pathogen reduction (for PC)	Sepsis: PC: 5–30; with bacterial detection: 2–20; with pathogen reduction: <0.5 RBCC: <0.2
	<i>Treponema pallidum</i> (syphilis)	~1 ^a	Serology ^{b,c}	<0.1
Virus	HIV-1/2	~0.1	Serology, NAT (+/- p24 Ag) ^{b,c}	0.1–1 ^d
	HBV	~0.5	Serology, NAT ^{b,c}	<0.5 (3 without NAT) ^d
	HCV	0.2–1.2	Serology, NAT ^{b,c}	<0.1–1 ^d
	HTLV-1/2	0.05–0.1 ^a	Serology, BC de leukocytation ^{b,c}	0.1–0.3 ^d
	HEV	0–10 (in endemic regions)	NAT	Endemic regions: <0.1 with NAT; a transmission rate from infected donors of ~50% has been reported
	CMV	Undetermined	Serology, BC de leukocytation ^{b,c}	<0.1 in de leukocyted BCs
	Parvovirus B19	~0.5 with viral DNA >10 ⁶ IU/mL, ^e up to 100 overall	NAT	Most adults are immune to parvovirus B19; up to 0.12% in seronegative adults has been reported
	West Nile virus	Up to 3 in high season endemic regions ^a	NAT ^b	High season endemic regions: <1 with NAT
Parasite	<i>Plasmodium</i> (Malaria)	~4 (40–50 in donors from endemic regions) ^a	Serology (NAT may be soon available)	<0.1 in non endemic regions
	<i>Babesia</i>	~90 (in endemic regions) ^a	Serology (NAT implementation is underway)	ND (0.04% donors may be within the serology window period)
	<i>Trypanosoma cruzi</i> (Chagas disease)	~0.14 in donors/mothers from endemic regions ^a	Serology	ND

^aAs assessed based on seropositivity, i.e., including a varying percentage of individuals not harboring the pathogen in their blood. ^bPrevention measures may also include pathogen reduction (for PC and plasma). ^cPrevention measures may also include a quarantine of the (cryopreserved) BC pending a negative serology on a subsequent donation (for plasma). ^dEstimated residual risk. ^eTransfusion risk deemed as absent below this threshold.

Note. Other pathogens associated with transfusion-transmitted infections at a very low frequency include arboviruses other than West Nile (dengue, Zika virus), hepatitis A, human herpesvirus-8, Japanese encephalitis virus, tick-borne encephalitis virus complex, and the prion responsible for variant Creutzfeldt-Jakob disease (4 cases in the United Kingdom, in the context of the bovine spongiform encephalopathy epidemic, before implementation of systematic de leukocytation).

Abbreviations: Ag, antigen; BC, blood component; CMV, cytomegalovirus; HBV, hepatitis B virus; HCV, hepatitis C virus; HEV, hepatitis E virus; HTLV, human T-cell leukemia virus; NAT, nucleic acid detection test; ND, not determined; PC platelet concentrate; RBCC, red blood cell concentrate.

at 4°C and therefore may be implicated in infections related to RBCC transfusion. Recipients of contaminated BCs may develop abrupt (during transfusion and up to several hours after) fever and chills, which can deteriorate to septic shock, DIC, and death. Endotoxin formed within the BC may be implicated. After sampling for bacterial culture, broad-spectrum antibiotics should be promptly initiated.

Pathogen reduction of platelets and plasma, and perhaps soon of RBCs as well, offers an additional means of reducing transfusion infection risks. Although effective for a wide range of pathogens, such processes are most often ineffective for bacterial spores and nonenveloped viruses such as hepatitis A virus (HAV), parvovirus B19, and hepatitis E virus (HEV). Postdonation information provided by the donor (i.e., fever occurring within 24 h after donation) may allow the involved blood products to be quarantined and provide an additional safety measure.

Transfusion-transmitted infections are increasingly rare. However, new or previously unidentified infectious risks may occur, as highlighted by the emergence of the transfusion-associated West Nile virus infection and babesiosis in early 2000 in the United States, as well as transfusion-associated hepatitis E in early 2010 in Europe. Such occurrences require active surveillance programs and the appropriate implementation of mitigation measures such as additional testing, pathogen reduction, and travel-related deferral criteria. Along with West Nile virus, a number of other arbovirus-related infections possibly transmissible by blood transfusion are endemic or involved in large epidemic outbreaks. Despite being possibly present in the blood at asymptomatic phases of the disease, documented cases of transfusion-transmitted infections involving these arboviruses have been very rare (Zika), without a discernible clinical impact (Dengue), or absent (Chikungunya). Route of infection (i.e., intravenous vs mosquito bite), pathogen dose, ability to

survive in the BC, storage temperature and duration, recipient immune status, and ongoing treatments may all impact the ability of a pathogen in the donor to induce a disease in the recipient. Estimated frequencies of transfusion-relevant infections in donors and of transfusion-transmitted infections are reported in Table 113-6. Such frequencies depend heavily on variables such as local epidemiology, donor deferral rules, risk reduction measures, and data reporting, and may vary considerably.

ALTERNATIVES AND PERSPECTIVES

In addition to promoting appropriate transfusion indications, patient blood management programs have highlighted a number of transfusion-sparing strategies, such as the treatment of anemia and/or iron deficiency before surgery, minimization of blood loss, and optimization of patient red cell mass. Erythropoietin stimulates erythrocyte production in patients with anemia from chronic renal failure and other conditions, thus avoiding or reducing the need for transfusion. Thrombopoietin receptor agonists has been shown to reduce platelet transfusion needs resulting from chemotherapy-induced thrombopenia. Gene therapy approaches in patients with sickle cell or major thalassemia offer the potential of dramatically reducing their transfusion needs. Stem cell-derived blood cells such as RBCs or platelets may in the future become a suitable alternative to rare blood donors.

Importantly, issues surrounding transfusion safety have evolved significantly and now fully encompass transfusion efficacy. New means of assessing transfusion efficacy are needed. Large-scale biological and population-based databases pertaining to blood donors and transfused patients will also be instrumental in assessing and understanding the basis of transfusion efficacy. Optimal transfusion care may soon require consideration of new criteria in relation to donor, blood product, and/or recipient characteristics.

ACKNOWLEDGMENTS

The authors are indebted to Jeffery S. Dzieckowski and Kenneth C. Anderson, who co-authored the chapter in the previous edition and expertly paved the way for this chapter.

FURTHER READING

- CARSON JL et al: Indications for and adverse effects of red-cell transfusion. *N Engl J Med* 377:1261, 2017.
- DELANEY M et al: Transfusion reactions: Prevention, diagnosis, and treatment. *Lancet* 388:2825, 2016.
- PANCH SR et al: Hemolytic transfusion reactions. *N Engl J Med* 381:150, 2019.

can survive freezing and thawing with little, if any, damage, making it possible to remove and store a portion of the patient's own bone marrow for later reinfusion following treatment of the patient with high-dose myelotoxic therapy.

CATEGORIES OF HEMATOPOIETIC CELL TRANSPLANTATION

Hematopoietic cell transplantation can be described according to the relationship between the patient and the donor and by the anatomic source of stem cells. In ~1% of cases, patients have identical twins who can serve as donors. With the use of syngeneic donors, there is no risk of graft-versus-host disease (GVHD), and unlike the use of autologous marrow, there is no risk that the stem cells are contaminated with tumor cells.

Allogeneic transplantation involves a donor and a recipient who are not genetically identical. Following allogeneic transplantation, immune cells transplanted with the stem cells or developing from them can react against the patient, causing GVHD. Alternatively, if the immunosuppressive preparative regimen used to treat the patient before transplant is inadequate, immunocompetent cells of the patient can cause graft rejection. The risks of these complications are greatly influenced by the degree of matching between donor and recipient for human leukocyte antigen (HLA) molecules encoded by genes of the major histocompatibility complex.

HLA molecules are responsible for binding antigenic proteins and presenting them to T cells. The antigens presented by HLA molecules may derive from exogenous sources (e.g., during active infections) or may be endogenous proteins. If individuals are not HLA-matched, T cells from one individual will react strongly to the mismatched HLA, or "major antigens," of the second. Even if the individuals are HLA-matched, the T cells of the donor may react to differing endogenous or "minor antigens" presented by the HLA of the recipient. Reactions to minor antigens tend to be less vigorous. The genes of major relevance to transplantation include HLA-A, -B, -C, and -D; they are closely linked and therefore tend to be inherited as haplotypes, with only rare crossovers between them. Thus, the odds that any one full sibling will match a patient are one in four, and the probability that the patient has an HLA-identical sibling is $1 - (0.75)n$, where n equals the number of siblings.

With conventional techniques, the risk of graft rejection is 1–3%, and the risk of severe, life-threatening acute GVHD is ~15% following transplantation between HLA-identical siblings. The incidence of graft rejection and GVHD increases progressively with the use of family member donors mismatched for one, two, or three antigens. Although survival following a one-antigen mismatched transplant is not markedly altered, survival following two- or three-antigen mismatched transplants is reduced. Newer approaches to GVHD prophylaxis, including the use of posttransplant high-dose cyclophosphamide, make transplantation between donor/recipient pairs who share only one HLA haplotype possible. Since the formation of the National Marrow Donor Program and other registries, HLA-matched unrelated donors can be identified for many patients. The genes encoding HLA antigens are highly polymorphic, and thus the odds of any two unrelated individuals being HLA identical are extremely low, somewhat less than 1 in 10,000. However, by recruiting >30 million volunteer donors, HLA-matched donors can be found for ~60% of patients for whom a search is initiated, with higher rates among whites and lower rates among minorities and patients of mixed race. It takes, on average, 3–4 months to complete a search and schedule and initiate an unrelated donor transplant. With improvements in HLA typing and supportive care measures, survival following matched unrelated donor transplantation is essentially the same as that seen with HLA-matched siblings.

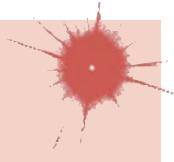
Allogeneic hematopoietic cell transplantation can be carried out across ABO blood barriers by removing isoagglutinins and/or incompatible red blood cells from the donor graft. However, depending on the direction of the mismatch, hemolysis of donor cells by persistent isoagglutinins in the host, or hemolysis of recipient red cells by isoagglutinins in the graft or developing from it may occur despite appropriate manipulation of the donor cell product.

Autologous transplantation involves the removal and storage of the patient's own stem cells with subsequent reinfusion after the patient

114

Hematopoietic Cell Transplantation

Frederick R. Appelbaum



Bone marrow transplantation was the original term used to describe the collection and transplantation of hematopoietic stem cells, but with the demonstration that peripheral blood and umbilical cord blood are also useful sources of stem cells, *hematopoietic cell transplantation* has become the preferred generic term for this process. Hematopoietic cell transplantation is used to treat patients with an abnormal but nonmalignant lymphohematopoietic system by replacing it with one from a normal donor. Hematopoietic cell transplantation is also used to treat malignancy by allowing the administration of higher doses of myelosuppressive therapy than would otherwise be possible and, in the setting of allogeneic Hematopoietic cell transplantation, by conferring an immunologic graft-versus-tumor effect. The use of hematopoietic cell transplantation is increasing, as it becomes safer and applicable to more diseases and as donor availability expands.

The Center for International Blood and Marrow Transplant Research (<http://www.cibmtr.org>) estimates that worldwide about 100,000 transplants were performed in 2020. The frequency of transplantation varied widely from country to country, with a close association of transplant rates with gross national income (GNI) per capita. However, even among countries with similar GNIs per capita, there are substantial differences between countries and regions regarding the frequency of transplantation, disease indications, and choice of donor type.

THE HEMATOPOIETIC STEM CELL

Several features of the hematopoietic stem cell make transplantation clinically feasible, including its remarkable regenerative capacity, its ability to home to the marrow space following intravenous injection, and the ability of the stem cell to be cryopreserved (Chap. 96). Transplantation of a single stem cell can replace the entire lymphohematopoietic system of an adult mouse. In humans, transplantation of a small percentage of a donor's bone marrow volume regularly results in complete and sustained replacement of the recipient's entire lymphohematopoietic system, including all red cells, granulocytes, B and T lymphocytes, and platelets, as well as cells comprising the fixed macrophage population, including Kupffer cells of the liver, pulmonary alveolar macrophages, osteoclasts, and Langerhans cells of the skin. The ability of the hematopoietic stem cell to home to the marrow following intravenous injection is mediated, in part, by an interaction between CXCL12, also known as stromal cell-derived factor 1, produced by marrow stromal cells and the alpha-chemokine receptor CXCR4 found on stem cells. Homing is also influenced by the interaction of cell-surface molecules, termed *selectins*, including E- and L-selectin, on bone marrow endothelial cells with ligands, termed *integrins*, such as VLA-4, on early hematopoietic cells. Human hematopoietic stem cells

receives high-dose myeloablative therapy. Unlike allogeneic transplantation, there is no risk of GVHD or graft rejection with autologous transplantation. On the other hand, autologous transplantation lacks a graft-versus-tumor (GVT) effect, and the autologous stem cell product can be contaminated with tumor cells, which could lead to relapse. A variety of techniques have been developed to “purge” autologous products of tumor cells, but no prospective randomized trials have shown that any approach decreases relapse rates or improves disease-free or overall survival.

Bone marrow aspirated from the posterior and anterior iliac crests initially was the source of hematopoietic stem cells for transplantation. Typically, anywhere from 1.5 to 5×10^8 nucleated marrow cells per kilogram are collected for allogeneic transplantation. Several studies have found improved survival following both matched sibling and unrelated transplantation by transplanting higher numbers of bone marrow cells.

Hematopoietic stem cells circulate in the peripheral blood but in very low concentrations. Following the administration of a myeloid growth factor such as granulocyte colony-stimulating factor (G-CSF) and during recovery from intensive chemotherapy, the concentration of hematopoietic progenitor cells in blood, as measured either by colony-forming units or expression of the CD34 antigen, increases markedly. This makes it possible to harvest adequate numbers of stem cells from the peripheral blood for transplantation. Donors are typically treated with 4 or 5 days of hematopoietic growth factor, following which stem cells are collected in one or two 4-h pheresis sessions. In the autologous setting, transplantation of $>2.5 \times 10^6$ CD34 cells per kilogram, a number that can be collected in most circumstances, leads to rapid and sustained engraftment in virtually all cases. In the 5–10% of patients who fail to mobilize enough CD34+ cells with growth factor alone, the addition of plerixafor, an antagonist of CXCR4, may be useful. Blocking CXCR4 allows more stem cells to escape the marrow. When compared to the use of autologous marrow, use of peripheral blood stem cells results in more rapid hematopoietic recovery. Although this more rapid recovery diminishes the morbidity rate of transplantation, no studies show improved survival.

In the setting of allogeneic transplantation, the use of growth factor-mobilized peripheral blood stem cells also results in faster engraftment than seen with marrow but at the cost of more chronic GVHD because of donor T-cell contamination. With matched sibling donors, the increased chronic GVHD is more than balanced by reductions in relapse rates and nonrelapse mortality rates, resulting in improved overall survival. However, in the setting of matched unrelated donor transplantation, use of peripheral blood results in more chronic GVHD without a compensatory survival advantage, favoring the use of bone marrow.

Umbilical cord blood contains a high concentration of hematopoietic progenitor cells, allowing for its use as a source of stem cells for transplantation. Cord blood transplantation from family members has been used when the immediate need for transplantation precludes waiting the 9 or so months generally required for the baby to mature to the point of donating marrow. Use of cord blood results in slower peripheral count recovery than seen with marrow but a lower incidence of GVHD, perhaps reflecting the low number of T cells in cord blood. Multiple cord blood banks have been developed to harvest and store cord blood for possible transplantation to unrelated patients from material that would otherwise be discarded. Currently $>800,000$ units are cryopreserved and available for use. The advantages of unrelated cord blood are rapid availability and decreased immune reactivity allowing for the use of partially matched units, which is of particular importance for those without matched unrelated donors. The risks of graft failure and transplant-related mortality are related to the dose of cord blood cells per kilogram, which previously limited the application of single cord blood transplantation to pediatric and smaller adult patients. Subsequent trials have found that for patients without suitable single cord units, the use of double cord transplants diminishes the risk of graft failure and early mortality even though only one of the donors ultimately engrafts. Given the similar survival rates seen with cord blood, matched unrelated, and haploidentical family member donors, a source of allogeneic stem cells can now be found for almost every patient in need (**Table 114-1**).

TABLE 114-1 Probability of Identifying a Donor Based on Stem Cell Source and Patient Ethnicity

	UNRELATED ADULT %	UNRELATED CORD %	HAPLOIDENTICAL
Ethnicity	8/8 ^a	7/8 ^a	≥4/6 ^b
Caucasian	75	90	>95
Hispanic	35	75	95
Black	18	70	90
			95

^aMatching for HLA-A, -B, -C, and DRB1. ^bMatching for HLA-A, -B, and DRB1.

THE TRANSPLANT PREPARATIVE REGIMEN

The treatment regimen administered to patients immediately preceding transplantation is designed to eradicate the patient's underlying disease and, in the setting of allogeneic transplantation, immunosuppress the patient adequately to prevent rejection of the transplanted stem cells. The appropriate regimen therefore depends on the disease setting and graft source. For example, when transplantation is performed to treat severe combined immunodeficiency and the donor is a histocompatible sibling, no treatment is needed because no host cells require eradication and the patient is already too immune-incompetent to reject the transplanted graft. For aplastic anemia, there is no large population of cells to eradicate, and high-dose cyclophosphamide plus antithymocyte globulin are sufficient to immunosuppress the patient adequately to accept the marrow graft. In the setting of thalassemia and sickle cell anemia, high-dose busulfan is frequently added to cyclophosphamide to eradicate hyperplastic host hematopoiesis. A variety of different regimens have been developed to treat malignant diseases. Most regimens include agents with high activity against the tumor in question at conventional doses and with myelosuppression as their predominant dose-limiting toxicity. Therefore, these regimens commonly include busulfan, cyclophosphamide, melphalan, thioguanine, etoposide, and total-body irradiation in various combinations.

Although high-dose treatment regimens were the initial approach to transplantation for malignancies, the realization that much of the antitumor effect of transplantation derives from an immunologically mediated GVT response led investigators to ask if reduced-intensity conditioning regimens might be effective and more tolerable. Evidence for a GVT effect comes from studies showing that posttransplant relapse rates are lowest in patients who develop acute and chronic GVHD, higher in those without GVHD, and higher still in recipients of T cell-depleted allogeneic or syngeneic marrow. The demonstration that complete remissions can be obtained in many patients who have relapsed after transplant by simply administering viable lymphocytes from the original donor further strengthens the argument for a potent GVT effect. Accordingly, a variety of alternative regimens have been studied, ranging from nonmyeloablative, which are the very minimum required to achieve engraftment (e.g., fludarabine plus 200 cGy total-body irradiation) and would cause only transient myelosuppression if no transplant were performed, to so-called reduced-intensity regimens, which would cause significant but not necessarily fatal myelosuppression in the absence of transplantation (e.g., fludarabine plus melphalan). Studies to date document that engraftment can be readily achieved with less toxicity than seen with conventional transplantation. Complete sustained responses have been documented in many patients, particularly those with more indolent hematologic malignancies. In general, relapse rates are higher following reduced-intensity conditioning, but transplant-related mortality is lower, favoring the use of reduced-intensity conditioning in patients with significant comorbidities. High-dose regimens are favored in those felt able to tolerate the treatment, particularly if patients have any evidence of measurable disease at the time of transplantation.

THE TRANSPLANT PROCEDURE

Marrow is usually collected from the donor's posterior and sometimes anterior iliac crests, with the donor under general or spinal anesthesia. Typically, 10–15 mL/kg of marrow is aspirated, placed in heparinized media, and filtered through 0.3- and 0.2-mm screens to remove fat and bony spicules. The collected marrow may undergo further processing

depending on the clinical situation, such as the removal of red cells to prevent hemolysis in ABO-incompatible transplants, the removal of donor T cells to prevent GVHD, or attempts to remove possible contaminating tumor cells in autologous transplantation. Marrow donation is safe, with only very rare complications reported.

Peripheral blood stem cells are collected by leukapheresis after the donor has been treated with hematopoietic growth factors or, in the setting of autologous transplantation, sometimes after treatment with a combination of chemotherapy and growth factors. Stem cells for transplantation are infused through a large-bore central venous catheter. Such infusions are usually well tolerated, although occasionally patients develop fever, cough, or shortness of breath. These symptoms typically resolve with slowing of the infusion. When the stem cell product has been cryopreserved using dimethyl sulfoxide, patients sometimes experience short-lived nausea or vomiting due to the taste (and smell) of the cryoprotectant.

■ ENGRAFTMENT AND IMMUNE RECONSTITUTION

Peripheral blood counts reach their nadir several days to a week after transplant as a consequence of the preparative regimen; then cells produced by the transplanted stem cells begin to appear in the peripheral blood. The rate of recovery depends on the source of stem cells and use of posttransplant growth factors. If marrow is the source, recovery to 100 granulocytes/ μ L occurs on average by day 16 and to 500/ μ L by day 22. Use of G-CSF-mobilized peripheral blood stem cells speeds the rate of recovery by ~1 week compared to marrow, whereas engraftment following cord blood transplantation is typically delayed by ~1 week. Use of a myeloid growth factor after transplant accelerates recovery by 3–5 days. Platelet counts usually recover shortly after granulocytes.

While granulocytes and other components of innate immunity recover rapidly after hematopoietic cell transplantation, adaptive immunity, which consists of cellular (T cell) and humoral (B cell) immunity, may take 1–2 years to fully recover. Survival and peripheral expansion of infused donor T cells is the dominant mechanism for T cell recovery in the first months after hematopoietic cell transplantation and results in mostly CD8+ T cells with a limited repertoire. After several months, de novo generation of donor derived CD4+ and CD8+ T cells becomes dominant providing a more diverse T-cell repertoire. B-cell counts recover by 6 months after autologous hematopoietic cell transplantation and 9 months after allogeneic hematopoietic cell transplantation. In general, immune recovery occurs more rapidly after autologous than allogeneic hematopoietic cell transplantation and after receipt of unmodified grafts compared to the setting of in vivo or ex vivo T-cell depletion.

Following allogeneic transplantation, engraftment can be documented using fluorescence in situ hybridization of sex chromosomes if donor and recipient are sex-mismatched or by analysis of short tandem repeat polymorphisms after DNA amplification.

■ COMPLICATIONS FOLLOWING HEMATOPOIETIC CELL TRANSPLANTATION

Early Direct Chemoradiotoxicities The transplant preparative regimen may cause a spectrum of acute toxicities that vary according to intensity of the regimen and the specific agents used but frequently include nausea, vomiting, and mild skin erythema (Fig. 114-1). High-dose cyclophosphamide can result in hemorrhagic cystitis, which can usually be prevented by bladder irrigation or with the sulphydryl compound mercaptoethanesulfonate (MESNA). Most high-dose preparative regimens will result in oral mucositis, which typically develops 5–7 days after transplant and often requires narcotic analgesia. Use of a patient-controlled analgesic pump provides the greatest patient satisfaction and results in a lower cumulative dose of narcotic. Keratinocyte growth factor (palifermin) can shorten the duration of mucositis by several days following autologous transplantation. Patients begin losing their hair 5–6 days after transplant and by 1 week are usually profoundly pancytopenic.

Depending on the intensity of the conditioning regimen, 3–10% of patients will develop sinusoidal obstruction syndrome (SOS) of the liver (formerly called venoocclusive disease), a syndrome that results from direct cytotoxic injury to hepatic-venular and sinusoidal endothelium, with subsequent deposition of fibrin and the development of

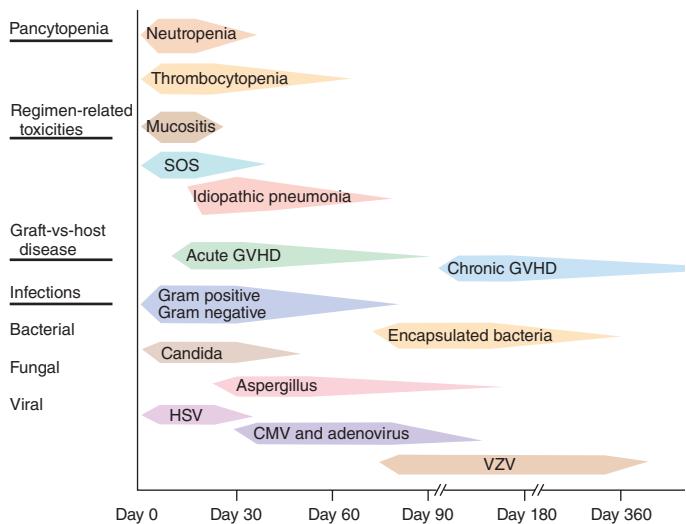


FIGURE 114-1 Major syndromes complicating marrow transplantation. CMV, cytomegalovirus; GVHD, graft-versus-host disease; HSV, herpes simplex virus; SOS, sinusoidal obstructive syndrome (formerly venoocclusive disease); VZV, varicella-zoster virus. The size of the shaded area roughly reflects the period of risk of the complication.

a local hypercoagulable state. This chain of events leads to the clinical symptoms of tender hepatomegaly, ascites, jaundice, and fluid retention. These symptoms can develop any time during the first month after transplant, with the peak incidence at day 16. Predisposing factors include prior exposure to intensive chemotherapy, pretransplant hepatitis of any cause, and use of more intense conditioning regimens. The mortality rate of sinusoidal obstruction syndrome is ~30%, with progressive hepatic failure culminating in a terminal hepatorenal syndrome. Treatment of severe SOS with defibrotide, a polydeoxyribonucleotide, reduces mortality.

Although most pneumonias developing early after transplant are caused by infectious agents, in a small percentage of patients, a diffuse interstitial pneumonia will develop that is a result of direct toxicity of high-dose preparative regimens. Bronchoalveolar lavage usually shows alveolar hemorrhage, and biopsies are typically characterized by diffuse alveolar damage, although some cases may have a more clearly interstitial pattern. High-dose glucocorticoids or antitumor necrosis factor therapies are sometimes used as treatment, although randomized trials proving their utility have not been reported.

Transplant-associated thrombotic microangiopathy is seen in 5–10% of patients, appearing on average about 1 month after transplant. The syndrome is characterized by presence of schistocytes on peripheral smear, elevated lactate dehydrogenase, thrombocytopenia, and acute kidney injury and is the result of endothelial injury and complement activation. Since calcineurin inhibitors are thought to contribute to the pathogenesis of the syndrome, changing immunosuppressive regimens is sometimes effective. Patients sometimes respond to eculizumab.

Late Direct Chemoradiotoxicities Two categories of chronic pulmonary disease occur in patients >3 months after hematopoietic cell transplantation. Cryptogenic organizing pneumonia is a restrictive lung disease characterized by dry cough, shortness of breath, and chest imaging showing a diffuse, fluffy infiltrate. Biopsy shows granulation tissue within alveolar spaces and small airways and no infectious agents. The disease responds well to corticosteroids and is entirely reversible. Bronchiolitis obliterans is an obstructive disease presenting with cough, progressive dyspnea, and radiologic evidence of air trapping. Pathology shows collagen and granulation tissue in and around bronchial structures and eventually obliteration of small airways. The disease is usually associated with chronic GVHD, and although it may respond to increasing immunosuppression, complete reversal is uncommon.

Other late complications of the preparative regimen include decreased growth velocity in children and delayed development of secondary sex characteristics. These complications can be partly ameliorated with the use of appropriate growth and sex hormone

replacement. Most men become azoospermic, and most postpubertal women will develop ovarian failure, which should be treated. However, pregnancy is possible after transplantation, and patients should be counseled accordingly. Thyroid dysfunction, usually well compensated, is sometimes seen. Cataracts develop in 10–20% of patients and are most common in patients treated with total-body irradiation and those who receive glucocorticoid therapy after transplant for treatment of GVHD. Aseptic necrosis of the femoral head is seen in 10% of patients and is particularly frequent following chronic glucocorticoid therapy. Both acute and late chemoradiotoxicities (except those due to glucocorticoids and other agents used to treat GVHD) are less frequent in recipients of reduced-intensity compared to high-dose preparative regimens.

Graft Failure Although complete and sustained engraftment is usually seen after transplant, occasionally marrow function either does not return or, after a brief period of engraftment, is lost. Graft failure after autologous transplantation can be the result of inadequate numbers of stem cells being transplanted, damage during ex vivo treatment or storage, or exposure of the patient to myelotoxic agents after transplant. Infections with cytomegalovirus (CMV) or human herpesvirus type 6 have also been associated with loss of marrow function. Graft failure after allogeneic transplantation can also be due to immunologic rejection of the graft by immunocompetent host cells. Such rejection is generally thought to be mostly T-cell mediated, but the presence pre-hematopoietic cell transplantation of donor-specific HLA antibodies in the patient is associated with poor engraftment, leading to the recommendation for screening for donor-directed anti-HLA antibodies in recipients prior to transplant. Immunologically based graft rejection is more common following use of less immunosuppressive preparative regimens, in recipients of T cell-depleted stem cell products, and in patients receiving grafts from HLA-mismatched donors or cord blood.

Treatment of graft failure involves removing all potentially myelotoxic agents from the patient's regimen and attempting a short trial of a myeloid growth factor. Persistence of lymphocytes of host origin in allogeneic transplant recipients with graft failure indicates immunologic rejection. Reinfusion of donor stem cells in such patients is usually unsuccessful unless preceded by a second immunosuppressive preparative regimen. Standard high-dose preparative regimens are tolerated poorly if administered within 100 days of a first transplant because of cumulative toxicities. However, reduced-intensity conditioning regimens have been effective in some cases.

Graft-Versus-Host Disease Acute GVHD occurs within the first 3 months after allogeneic transplant with a peak onset around 4 weeks and is characterized by an erythematous maculopapular rash; by persistent anorexia or diarrhea, or both; and by liver disease with increased serum levels of bilirubin, alanine and aspartate aminotransferase, and alkaline phosphatase. Because many conditions can mimic acute GVHD, the diagnosis usually requires skin, liver, or endoscopic biopsy for confirmation. In all these organs, endothelial damage and lymphocytic infiltrates are seen. In skin, the epidermis and hair follicles are damaged; in liver, the small bile ducts show segmental disruption; and in intestines, destruction of the crypts and mucosal ulceration may be noted. A commonly used rating system for acute GVHD is shown in **Table 114-2**. Grade I acute GVHD is of little clinical significance, does

not affect the likelihood of survival, and does not require treatment. In contrast, grades II to IV GVHD are associated with significant symptoms and a poorer probability of survival and require aggressive therapy. The incidence of acute GVHD is higher in recipients of stem cells from mismatched or unrelated donors, in older patients, and in patients unable to receive full doses of drugs used to prevent the disease.

Currently, the standard approach to GVHD prevention is the administration of a calcineurin inhibitor (cyclosporine or tacrolimus) combined with an antimetabolite (methotrexate or mycophenolate mofetil) following transplantation. The addition of anti-T-cell immune globulin (ATG) may further reduce the incidence of GVHD but has not been shown to improve survival. Other approaches being tested in phase 3 studies include the addition of sirolimus to the standard two-drug regimen, the removal of subsets or all T cells from the stem cell inoculum, and the use of cyclophosphamide administered several days after transplant in an effort to deplete activated alloreactive T cells.

Despite prophylaxis, significant acute GVHD will develop in ~30% of recipients of stem cells from matched siblings. Factors associated with a greater risk of acute GVHD include HLA-mismatching between recipient and donor, patient and donor age, use of more intense preparative regimens, and use of multiparous women as donors. Presumably, multiparous women have more alloreactivity based on carriage of genetically disparate fetuses. Disruption of the intestinal microbiota leading to loss of diversity and overgrowth by a single taxon is associated with a higher risk of GVHD and transplant-associated mortality. Biomarkers, including ST2, REG32, and TNF R1, have been identified that predict the severity of acute GVHD. The disease is usually treated with prednisone at a daily dose of 1–2 mg/kg. Patients in whom the acute GVHD fails to respond to prednisone sometimes respond to the oral JAK2 inhibitor ruxolitinib.

Chronic GVHD occurs most commonly between 3 months and 2 years after allogeneic transplant, developing in 20–50% of recipients. The disease is more common in older patients, with the use of peripheral blood rather than marrow as the stem cell source, in recipients of mismatched or unrelated stem cells, and in those with a preceding episode of acute GVHD. The disease resembles an autoimmune disorder with malar rash, sicca syndrome, arthritis, obliterative bronchiolitis, and bile duct degeneration with cholestasis. Mild chronic GVHD can sometimes be managed using local therapies (topical glucocorticoids to skin and cyclosporine eye drops). More severe disease requires systemic therapy usually with prednisone alone or in combination with cyclosporine. Ibrutinib is sometimes effective in patients whose disease does not respond to initial therapy. Mortality rates from chronic GVHD average around 15%, but range from 5 to 50% depending on severity. In most patients, chronic GVHD resolves, but it may require 1–3 years of immunosuppressive treatment before these agents can be withdrawn without the disease recurring. Because patients with chronic GVHD are susceptible to significant infection, they should receive prophylactic trimethoprim-sulfamethoxazole, and all suspected infections should be investigated and treated aggressively.

Although onset before or after 3 months after transplant is often used to discriminate between acute and chronic GVHD, occasional patients will develop signs and symptoms of acute GVHD after 3 months (late-onset acute GVHD), whereas others will exhibit signs

TABLE 114-2 Clinical Staging and Grading of Acute Graft-versus-Host Disease

CLINICAL STAGE	SKIN	LIVER—BILIRUBIN, $\mu\text{mol/L}$ (mg/dL)	GUT
1	Rash <25% body surface	34–51 (2–3)	Diarrhea 500–1000 mL/d
2	Rash 25–50% body surface	51–103 (3–6)	Diarrhea 1000–1500 mL/d
3	Generalized erythroderma	103–257 (6–15)	Diarrhea >1500 mL/d
4	Desquamation and bullae	>257 (>15)	Ileus
OVERALL CLINICAL GRADE	SKIN STAGE	LIVER STAGE	GUT STAGE
I	1–2	0	0
II	1–3	1	1
III	1–3	2–3	2–3
IV	2–4	2–4	2–4

and symptoms of both acute and chronic GVHD (overlap syndrome). There are as yet no data to suggest that these patients should be treated differently than those with classic acute or chronic GVHD.

From 3 to 5% of patients will develop an autoimmune disorder following allogeneic hematopoietic cell transplantation, most commonly autoimmune hemolytic anemia or idiopathic thrombocytopenic purpura. Unrelated donor source and chronic GVHD are risk factors, but autoimmune disorders have been reported in patients with no obvious GVHD. Treatment is with prednisone, cyclosporine, or rituximab.

Infection Posttransplant patients, particularly recipients of allogeneic transplantation, require unique approaches to the problem of infection. Early after transplantation, patients are profoundly neutropenic, and because the risk of bacterial infection is so great, most centers place patients on broad-spectrum antibiotics once the granulocyte count falls to <500/ μ L. Prophylaxis against fungal infections reduces rates of infection and improves overall survival. Fluconazole is often used for patients with standard risk, while prophylaxis with mold active agents (voriconazole or posaconazole) should be considered for patients at higher risk, such as those with a prior fungal infection. Patients seropositive for herpes simplex should receive acyclovir prophylaxis. One approach to infection prophylaxis is shown in **Table 114-3**. Despite these prophylactic measures, most patients will develop fever and signs of infection after transplant. The management of patients who become febrile despite bacterial and fungal prophylaxis is a difficult challenge and is guided by individual aspects of the patient and by the institution's experience.

The general problem of infection in the immunocompromised host is discussed in **Chap. 143**.

Once patients engraft, the incidence of bacterial infection diminishes; however, patients, particularly allogeneic transplant recipients, remain at significant risk of infection. During the period from engraftment until about 3 months after transplant, the most common causes of infection are gram-positive bacteria, fungi (particularly *Aspergillus*), and viruses including CMV. CMV disease, which in the past was frequently seen and often fatal, can be prevented in seronegative patients transplanted from seronegative donors by the use of either seronegative blood products or products from which the white blood cells have been removed. In seropositive patients or patients transplanted from seropositive donors, either prophylaxis or preemptive therapy is used. Letermovir administered over the first 3 months after transplant is effective as prophylaxis. An alternative approach is to monitor blood of patients after transplant using polymerase chain reaction assays for viral DNA and to treat reactivation preemptively with ganciclovir before clinical disease develops. Foscarnet is effective for some patients who develop CMV antigenemia or infection despite the use of ganciclovir or who cannot tolerate the drug, but it can be associated with severe electrolyte wasting.

Pneumocystis jirovecii pneumonia, once seen in 5–10% of patients, can be prevented by treating patients with oral trimethoprim-sulfamethoxazole for 1 week before transplant and resuming the treatment once patients engraft.

TABLE 114-3 Approach to Infection Prophylaxis in Allogeneic Transplant Recipients

ORGANISM	AGENT	APPROACH
Bacterial	Levofloxacin	750 mg PO or IV daily
Fungal	Fluconazole	400 mg PO qd to day 75 posttransplant
<i>Pneumocystis jirovecii</i>	Trimethoprim-sulfamethoxazole	1 double-strength tablet PO bid 2 days/week until day 180 or off immunosuppression
Viral		
Herpes simplex	Acyclovir	800 mg PO bid to day 30
Varicella-zoster	Acyclovir	800 mg PO bid to day 365
Cytomegalovirus	Ganciclovir	5 mg/kg IV bid for 7 days, then 5 (mg/kg)/d 5 days/week to day 100

Respiratory viruses that cause community-acquired infections, including respiratory syncytial virus (RSV), parainfluenza virus, influenza virus, and metapneumovirus, can be life threatening or fatal in the post-transplant patient. Protection of patients from infected visitors and staff by avoiding such contacts is critical. Neuraminidase inhibitors are effective for influenza infections. Inhaled ribavirin is sometimes used for RSV.

The risk of infection diminishes considerably beyond 3 months after transplant unless chronic GVHD requiring continuous immunosuppression develops. Most transplant centers recommend continuing trimethoprim-sulfamethoxazole prophylaxis while patients are receiving any immunosuppressive drugs and also recommend careful monitoring for late CMV reactivation. In addition, many centers recommend prophylaxis against varicella-zoster, using acyclovir for 1 year after transplant. Patients should be revaccinated against tetanus, diphtheria, *Haemophilus influenzae*, polio, and pneumococcal pneumonia starting at 12 months after transplant and against measles, mumps, and rubella (MMR), varicella-zoster virus, and possibly pertussis at 24 months.

TREATMENT

Nonmalignant Diseases

Evidence-based indications for hematopoietic cell transplantation have been published by several organizations and are guided not only by disease-related factors but also by patient comorbidities, socioeconomic issues, caregiver and donor availability, and patient preference.

IMMUNODEFICIENCY DISORDERS

By replacing abnormal stem cells with cells from a normal donor, hematopoietic cell transplantation can cure patients of a variety of immunodeficiency disorders including severe combined immunodeficiency, Wiskott-Aldrich syndrome, and Chédiak-Higashi syndrome. The widest experience is with severe combined immunodeficiency disease, where cure rates of 90% can be expected with HLA-identical donors and success rates of 50–70% have been reported using haplotype-mismatched parents as donors (**Table 114-4**).

APLASTIC ANEMIA

Transplantation from matched siblings after a preparative regimen of high-dose cyclophosphamide and antithymocyte globulin cures up to 90% of patients age <40 years with severe aplastic anemia. Results in older patients and in recipients of mismatched family member or unrelated marrow are less favorable; therefore, a trial of immunosuppressive therapy is generally recommended for such patients before considering transplantation. Transplantation is effective in all forms of aplastic anemia including, for example, the syndromes associated with paroxysmal nocturnal hemoglobinuria and Fanconi's anemia. Patients with Fanconi's anemia are abnormally sensitive to the toxic effects of alkylating agents, and so less intensive preparative regimens are used in their treatment (**Chap. 102**).

HEMOGLOBINOPATHIES

Marrow transplantation from an HLA-identical sibling following a preparative regimen of busulfan and cyclophosphamide can cure 80–90% of patients with thalassemia major. The best outcomes can be expected if patients are transplanted before they develop hepatomegaly or portal fibrosis and if they have been given adequate iron chelation therapy. Among such patients, the probabilities of 5-year survival and disease-free survival are 95 and 90%, respectively. Although prolonged survival can be achieved with aggressive chelation therapy, transplantation is the only curative treatment for thalassemia. Transplantation is potentially curative for patients with sickle cell anemia. Two-year survival and disease-free survival rates of 95 and 85%, respectively, have been reported following matched sibling or cord blood transplantation. Decisions about patient selection and the timing of transplantation remain difficult, but transplantation is a reasonable option for children and young

TABLE 114-4 Estimated 5-Year Survival Rates Following Transplantation^a

DISEASE	ALLOGENEIC, %	AUTOLOGOUS, %
Severe combined immunodeficiency	90	N/A
Aplastic anemia	90	N/A
Thalassemia	90	N/A
Acute myeloid leukemia		
First remission	55–60	50
Second remission	40	30
Acute lymphocytic leukemia		
First remission	50	40
Second remission	40	30
Chronic myeloid leukemia		
Chronic phase	70	ID
Accelerated phase	40	ID
Blast crisis	15	ID
Chronic lymphocytic leukemia	50	ID
Myelodysplasia	45	ID
Multiple myeloma—initial therapy	N/A	60
Non-Hodgkin's lymphoma		
First relapse/second remission	40	40
Hodgkin's disease		
First relapse/second remission	40	50

^aThese estimates are generally based on data reported by the International Bone Marrow Transplant Registry. The analysis has not been reviewed by their Advisory Committee.

Abbreviations: ID, insufficient data; N/A, not applicable.

adults who have suffered complications of sickle cell anemia including stroke, recurrent vasoocclusive pain, sickle cell lung disease, or sickle nephropathy ([Chap. 98](#)).

OTHER NONMALIGNANT DISEASES

Theoretically, hematopoietic cell transplantation should be able to cure any disease that results from an inborn error of the lymphohematopoietic system. Transplantation has been used successfully to treat congenital disorders of white blood cells such as Kostmann's syndrome, chronic granulomatous disease, and leukocyte adhesion deficiency. Congenital anemias such as Blackfan-Diamond anemia can also be cured with transplantation. Since the penetrance of some congenital marrow failure states is variable, potential family member donors should be carefully screened before use to assure they are not affected. Infantile malignant osteopetrosis is due to an inability of the osteoclast to resorb bone, and because osteoclasts derive from the marrow, transplantation can cure this rare inherited disorder.

Hematopoietic cell transplantation has been used as treatment for a number of storage diseases caused by enzymatic deficiencies, such as Gaucher's disease, Hurler's syndrome, Hunter's syndrome, and infantile metachromatic leukodystrophy. Transplantation for these diseases has not been uniformly successful, but treatment early in the course of these diseases, before irreversible damage to extramedullary organs has occurred, increases the chance for success.

Transplantation is being explored as a treatment for severe acquired autoimmune disorders. These trials are based on studies demonstrating that transplantation can reverse autoimmune disorders in animal models and on the observation that occasional patients with coexisting autoimmune disorders and hematologic malignancies have been cured of both with transplantation. A prospective randomized trial found that patients with severe scleroderma have improved event-free and overall survival if treated with hematopoietic cell transplantation.

ACUTE LEUKEMIA

Allogeneic hematopoietic cell transplantation cures 15–20% of patients who do not achieve complete response after induction chemotherapy for acute myeloid leukemia (AML) and is the only form of therapy that can cure such patients. Thus, all patients with AML who are possible transplant candidates should have their HLA type determined soon after diagnosis to enable hematopoietic cell transplantation for those who fail to enter remission. Cure rates of 30–35% are seen when patients are transplanted in second remission or in first relapse. The best results with allogeneic transplantation are achieved when applied during first remission, with disease-free survival rates averaging 55–60%. Meta-analyses of studies comparing matched related donor transplantation to chemotherapy for adult AML patients age <60 years show a survival advantage with transplantation. This advantage is greatest for those with unfavorable-risk AML and is lost in those with favorable-risk disease. While hematopoietic cell transplantation can be performed in patients up to age 75 and possibly beyond, prospective trials comparing hematopoietic cell transplantation with chemotherapy are lacking for older patients. The role of autologous transplantation in the treatment of AML is less well defined. The rates of disease recurrence with autologous transplantation are higher than those seen after allogeneic transplantation, and cure rates are somewhat less.

Similar to patients with AML, adults with acute lymphocytic leukemia who do not achieve a complete response to induction chemotherapy can be cured in 15–20% of cases with immediate transplantation. Cure rates improve to 30–50% in second remission, and therefore, transplantation can be recommended for adults who have persistent disease after induction chemotherapy or who subsequently relapse. Transplantation in first remission results in cure rates of about 55%. Transplantation appears to offer a survival advantage over chemotherapy for patients with high-risk disease as defined by molecular profiling. Debate continues about whether adults with standard-risk disease should be transplanted in first remission or whether transplantation should be reserved until relapse. Autologous transplantation is associated with a higher relapse rate but a somewhat lower risk of nonrelapse mortality when compared to allogeneic transplantation. There is no obvious role of autologous transplantation for acute lymphocytic leukemia in first remission, and for second-remission patients, most experts recommend use of allogeneic stem cells if an appropriate donor is available.

CHRONIC LEUKEMIA

Allogeneic hematopoietic cell transplantation is indicated for patients with chronic myeloid leukemia (CML) who are in chronic phase but have failed therapy with two or more tyrosine kinase inhibitors. In such patients, cure rates of 70% can be expected. Hematopoietic cell transplantation is also recommended for patients with CML who present or progress to accelerated phase or blast crisis, although lower cure rates are seen in such patients ([Chap. 105](#)).

Although allogeneic transplantation can cure patients with chronic lymphocytic leukemia (CLL), it has not been extensively studied because of the chronic nature of the disease, the age profile of patients, and more recently, the availability of multiple effective therapies. In those cases where it was studied, complete remissions were achieved in the majority of patients, with disease-free survival rates of ~50% at 3 years, despite the advanced stage of the disease at the time of transplant.

MYELODYSPLASIA AND MYELOPROLIFERATIVE DISORDERS

Between 20 and 65% of patients with myelodysplasia appear to be cured with allogeneic transplantation. Results are better among younger patients and those with less advanced disease. However, patients with early-stage myelodysplasia can live for extended periods without intervention, and so transplantation is generally reserved for patients with an International Prognostic Scoring System (IPSS) score of Int-2 or higher, or for selected patients with an IPSS score of Int-1 who have other poor prognostic features

(Chap. 102). Allogeneic hematopoietic cell transplantation can cure patients with primary myelofibrosis or myelofibrosis secondary to polycythemia vera or essential thrombocythemia, with 5-year progression-free survival rates in excess of 65% being reported. It may require many months for the fibrosis to resolve.

LYMPHOMA

Patients with disseminated intermediate- or high-grade non-Hodgkin's lymphoma who have not been cured by first-line chemotherapy and are transplanted in first relapse or second remission can still be cured in 40–50% of cases. This represents a clear advantage over results obtained with conventional-dose salvage chemotherapy. It is unsettled whether patients with high-risk disease benefit from transplantation in first remission. Most experts favor the use of autologous rather than allogeneic transplantation for patients with intermediate- or high-grade non-Hodgkin's lymphoma, because fewer complications occur with this approach and survival appears equivalent. Although autologous transplantation results in high response rates in patients with recurrent disseminated indolent non-Hodgkin's lymphoma, the availability of newer agents for this category of patient leaves the role of transplantation unsettled. Reduced-intensity conditioning regimens followed by allogeneic transplantation result in high rates of complete and enduring complete responses in patients with recurrent indolent lymphomas.

The role of transplantation in Hodgkin's disease is similar to that in intermediate- and high-grade non-Hodgkin's lymphoma. With transplantation, 5-year disease-free survival is 20–30% in patients who never achieve a first remission with standard chemotherapy and up to 70% for those transplanted in second remission. Transplantation has no defined role in first remission in Hodgkin's disease.

MYELOMA

Patients with myeloma whose disease progresses after first-line therapy can sometimes benefit from allogeneic or autologous transplantation. Prospective randomized studies demonstrate that the inclusion of autologous transplantation as part of initial therapy results in improved disease-free survival and overall survival. Further benefit is seen with the use of lenalidomide maintenance therapy following transplantation. The use of autologous transplantation followed by nonmyeloablative allogeneic transplantation has yielded mixed results.

SOLID TUMORS

Patients with testicular cancer in whom first-line platinum-containing chemotherapy has failed can still be cured in ~50% of cases if treated with high-dose chemotherapy with autologous stem cell support, an outcome better than that seen with low-dose salvage chemotherapy. The use of high-dose chemotherapy with autologous stem cell support is being studied for several other solid tumors, including neuroblastoma and pediatric sarcomas. As in most other settings, the best results were obtained in patients with limited amounts of disease and in whom the remaining tumor remains sensitive to conventional-dose chemotherapy. Few randomized trials of transplantation in these diseases have been completed.

POSTTRANSPLANT RELAPSE

Patients who relapse following autologous transplantation sometimes respond to further chemotherapy and may be candidates for possible allogeneic transplantation, particularly if the remission following the initial autologous transplant was long. Several options are available for patients who relapse following allogeneic transplantation. Treatment with infusions of unirradiated donor lymphocytes results in complete responses in as many as 75% of patients with chronic myeloid leukemia, 40% with myelodysplasia, 25% with AML, and 15% with myeloma. Major complications of donor lymphocyte infusions include transient myelosuppression and the development of GVHD. These complications depend on the number of donor lymphocytes given and the schedule of infusions, with less GVHD seen with lower dose, fractionated schedules.

FURTHER READING

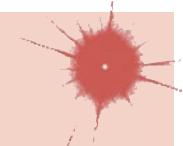
- HOURIGAN CS et al: Impact of conditioning intensity of allogeneic transplantation for acute myeloid leukemia with genomic evidence of residual disease. *J Clin Oncol* 38:1273, 2020.
- JAGASIA M et al: Ruxolitinib for the treatment of steroid-refractory acute GVHD (REACH1): A multicenter, open-label phase 2 trial. *Blood* 135:1739, 2020.
- MAJHAIL NS et al: Indications for autologous and allogeneic hematopoietic cell transplantation: Guidelines from the American Society for Blood and Marrow Transplantation. *Biol Blood Marrow Transplant* 21:1863, 2015.
- MARTY FM et al: Letermovir prophylaxis for cytomegalovirus in hematopoietic-cell transplantation. *N Engl J Med* 377:2433, 2017.
- MCDONALD GB et al: Survival, nonrelapse mortality, and relapse-related mortality after allogeneic hematopoietic cell transplantation: Comparing 2003–2007 versus 2013–2017 cohorts. *Ann Intern Med* 172:229, 2020.
- MIKLOS D et al: Ibrutinib for chronic graft-versus-host disease after failure of prior therapy. *Blood* 130:2243, 2017.
- PELED JU et al: Microbiota as predictor of mortality in allogeneic hematopoietic-cell transplantation. *N Engl J Med* 382:822, 2020.
- SULLIVAN KM et al: Myeloablative autologous stem-cell transplantation for severe scleroderma. *N Engl J Med* 378:35, 2018.
- ZEISER R, BLAZAR BR: Acute graft-versus-host disease - biologic process, prevention, and therapy. *N Engl J Med* 377:2167, 2017.
- ZEISER R, BLAZAR BR: Pathophysiology of chronic graft-versus-host disease and therapeutic targets. *N Engl J Med* 377:2565, 2017.

Section 3 Disorders of Hemostasis

115

Disorders of Platelets and Vessel Wall

Barbara A. Konkle



Hemostasis is a dynamic process in which the platelet and the blood vessel wall play key roles. Platelets are activated upon adhesion to von Willebrand factor (VWF) and collagen in the exposed subendothelium after injury. Platelet activation is also mediated through shear forces imposed by blood flow itself, particularly in areas where the vessel wall is diseased, and is also affected by the inflammatory state of the endothelium. The activated platelet surface provides the major physiologic site for coagulation factor activation, which results in further platelet activation and fibrin formation. Genetic and acquired influences on the platelet and vessel wall, as well as on the coagulation and fibrinolytic systems, determine whether normal hemostasis or bleeding or clotting symptoms will result.

THE PLATELET

Platelets are released from the megakaryocyte, likely under the influence of flow in the capillary sinuses. The normal blood platelet count is 150,000–450,000/ μ L. The major regulator of platelet production is the hormone thrombopoietin (TPO), which is synthesized in the liver and other organs. Synthesis is increased with inflammation and specifically by interleukin 6. TPO binds to its receptor on platelets and megakaryocytes, by which it is removed from the circulation. Thus, a reduction in platelet and megakaryocyte mass increases the level of TPO, which then stimulates platelet production. Platelets circulate with an average life span of 7–10 days. Approximately one-third of the platelets reside in the spleen, and this number increases in proportion to splenic size, although the platelet count rarely decreases to <40,000/ μ L as the spleen

enlarges. Platelets are physiologically very active, but are anucleate, and thus have limited capacity to synthesize new proteins.

Normal vascular endothelium contributes to preventing thrombosis by inhibiting platelet function (Chap. 65). When vascular endothelium is injured, these inhibitory effects are overcome, and platelets adhere to the exposed intimal surface primarily through VWF, a large multimeric protein present in both plasma and in the extracellular matrix of the subendothelial vessel wall. Platelet adhesion results in the generation of intracellular signals that lead to activation of the platelet glycoprotein (Gp) IIb/IIIa ($\alpha_{IIb}\beta_3$) receptor and resultant platelet aggregation.

Activated platelets undergo release of their granule contents, which include nucleotides, adhesive proteins, growth factors, and procoagulants that serve to promote platelet aggregation and blood clot formation and influence the environment of the forming clot. During platelet aggregation, additional platelets are recruited to the site of injury, leading to the formation of an occlusive platelet thrombus. The platelet plug is stabilized by the fibrin mesh that develops simultaneously as the product of the coagulation cascade.

THE VESSEL WALL

Endothelial cells line the surface of the entire circulatory tree, totaling $1\text{--}6 \times 10^{13}$ cells, enough to cover a surface area equivalent to about six tennis courts. The endothelium is physiologically active, controlling vascular permeability, flow of biologically active molecules and nutrients, blood cell interactions with the vessel wall, the inflammatory response, and angiogenesis.

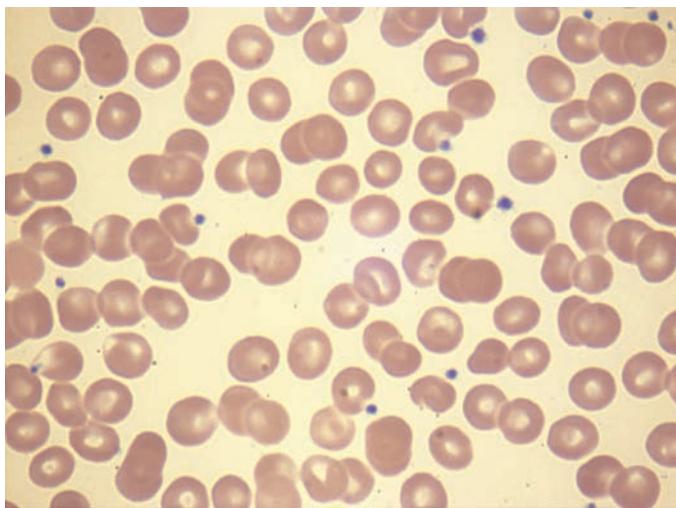
The endothelium normally presents an antithrombotic surface (Chap. 65) but rapidly becomes prothrombotic when stimulated,

which promotes coagulation, inhibits fibrinolysis, and activates platelets. In many cases, endothelium-derived vasodilators are also platelet inhibitors (e.g., nitric oxide), and conversely, endothelium-derived vasoconstrictors (e.g., endothelin) can also be platelet activators. The net effect of vasodilation and inhibition of platelet function is to promote blood fluidity, whereas the net effect of vasoconstriction and platelet activation is to promote thrombosis. Thus, blood fluidity and hemostasis are regulated by the balance of antithrombotic/prothrombotic and vasodilatory/vasoconstrictor properties of endothelial cells.

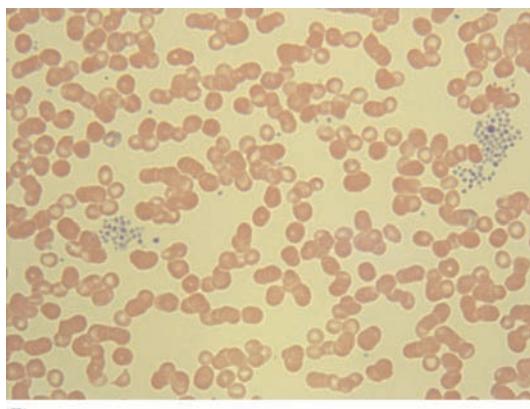
DISORDERS OF PLATELETS

■ THROMBOCYTOPENIA

Thrombocytopenia results from one or more of three processes: (1) decreased bone marrow production; (2) sequestration, usually in an enlarged spleen; and/or (3) increased platelet destruction. Disorders of production may be either inherited or acquired. In evaluating a patient with thrombocytopenia, a key step is to review the peripheral blood smear and to first rule out "pseudothrombocytopenia," particularly in a patient without an apparent cause for the thrombocytopenia. Pseudothrombocytopenia (Fig. 115-1B) is an in vitro artifact resulting from platelet agglutination via antibodies (usually IgG, but also IgM and IgA) when the calcium content is decreased by blood collection in ethylenediamine tetraacetic (EDTA) (the anticoagulant present in tubes [purple top] used to collect blood for complete blood counts [CBCs]). If a low platelet count is obtained in EDTA-anticoagulated blood, a blood smear should be evaluated and a platelet count determined in blood collected into sodium citrate (blue top tube) or heparin (green



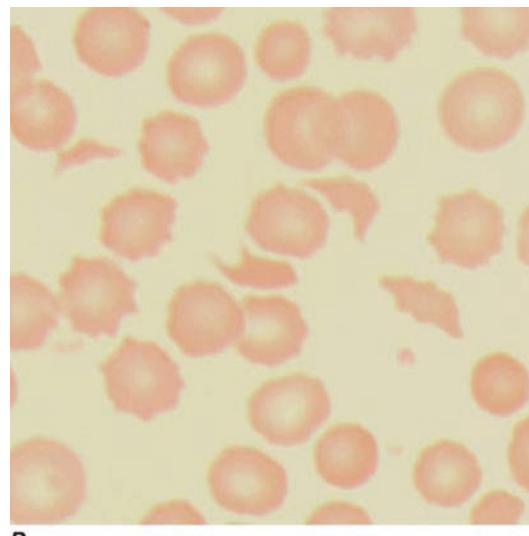
A



B



C



D

FIGURE 115-1 Photomicrographs of peripheral blood smears. A. Normal peripheral blood. **B.** Platelet clumping in pseudothrombocytopenia. **C.** Abnormal large platelet in autosomal dominant macrothrombocytopenia. **D.** Schistocytes and decreased platelets in microangiopathic hemolytic anemia.

top tube), or a smear of freshly obtained unanticoagulated blood, such as from a finger stick, can be examined.

APPROACH TO THE PATIENT

Thrombocytopenia

The history and physical examination, results of the CBC, and review of the peripheral blood smear are all critical components in the initial evaluation of thrombocytopenic patients (Fig. 115-2). The overall health of the patient and whether he or she is receiving drug treatment will influence the differential diagnosis. A healthy young adult with thrombocytopenia will have a much more limited differential diagnosis than an ill hospitalized patient who is receiving multiple medications. Except in unusual inherited disorders, decreased platelet production usually results from bone marrow disorders that also affect red blood cell (RBC) and/or white blood cell (WBC) production. Because myelodysplasia can present with isolated thrombocytopenia, the bone marrow should be examined in patients presenting with isolated thrombocytopenia who are older than 60 years of age or who do not respond to initial therapy. While inherited thrombocytopenia is rare, any prior platelet counts should be retrieved and a family history regarding thrombocytopenia obtained. A careful history of drug ingestion should be obtained, including nonprescription and herbal remedies, because drugs are the most common cause of thrombocytopenia.

The physical examination can document an enlarged spleen, evidence of chronic liver disease, and other underlying disorders. Mild to moderate splenomegaly may be difficult to appreciate in many individuals due to body habitus and/or obesity but can be easily assessed by abdominal ultrasound. A platelet count of approximately 5000–10,000 is required to maintain vascular integrity in the microcirculation. When the count is markedly decreased, petechiae first appear in areas of increased venous pressure, the ankles and feet in an ambulatory patient. Petechiae are pinpoint, nonblanching hemorrhages and are usually a sign of a decreased platelet number

and not platelet dysfunction. Wet purpura, blood blisters that form on the oral mucosa, are thought to denote an increased risk of life-threatening hemorrhage in the thrombocytopenic patient. Excessive bruising is seen in disorders of both platelet number and function.

Infection-Induced Thrombocytopenia Many viral and bacterial infections result in thrombocytopenia and are the most common noniatrogenic cause of thrombocytopenia. This may or may not be associated with laboratory evidence of disseminated intravascular coagulation (DIC), which is most commonly seen in patients with systemic infections with gram-negative bacteria and is seen in patients ill with COVID-19. Infections can affect both platelet production and platelet survival. In addition, immune mechanisms can be at work, as in infectious mononucleosis and early HIV infection. Late in HIV infection, pancytopenia and decreased and dysplastic platelet production are more common. Immune-mediated thrombocytopenia in children usually follows a viral infection and almost always resolves spontaneously. This association of infection with immune thrombocytopenic purpura is less clear in adults.

Drug-Induced Thrombocytopenia Many drugs have been associated with thrombocytopenia. A predictable decrease in platelet count occurs after treatment with many chemotherapeutic drugs due to bone marrow suppression (Chap. 73). Drugs that cause isolated thrombocytopenia and have been confirmed with positive laboratory testing are listed in Table 115-1, but all drugs should be suspect in a patient with thrombocytopenia without an apparent cause and should be stopped, or substituted, if possible. Although not as well studied, herbal and over-the-counter preparations may also result in thrombocytopenia and should be discontinued in patients who are thrombocytopenic.

Classic drug-dependent antibodies are antibodies that react with specific platelet surface antigens and result in thrombocytopenia only when the drug is present. Many drugs are capable of inducing these antibodies, but for some reason, they are more common with quinine and sulfonamides. Drug-dependent antibody binding can be demonstrated by laboratory assays, showing antibody binding in the presence of, but not without, the drug present in the assay. The thrombocytopenia typically occurs after a period of initial exposure (median length 21 days), or upon reexposure, and usually resolves in 7–10 days after drug withdrawal. The thrombocytopenia caused by the platelet Gp IIb/IIIa inhibitory drugs, such as abciximab, differs in that it may occur within 24 h of initial exposure. This appears to be due to the presence

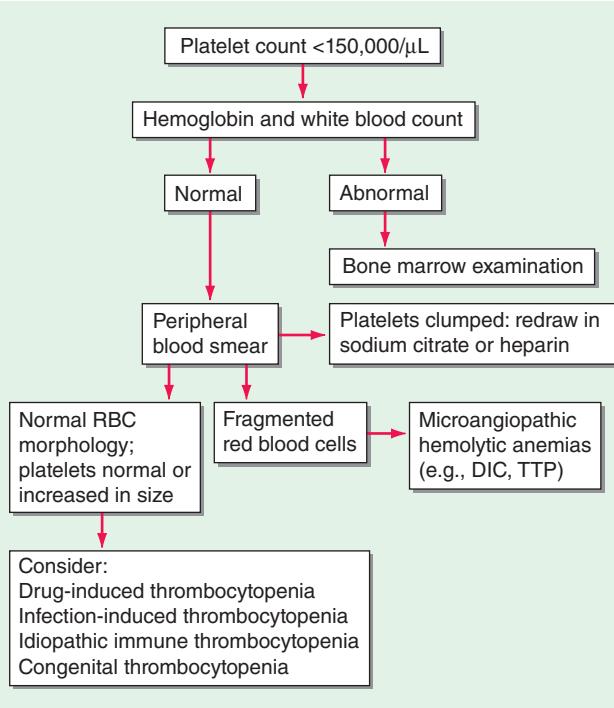


FIGURE 115-2 Algorithm for evaluating the thrombocytopenic patient. DIC, disseminated intravascular coagulation; RBC, red blood cell; TTP, thrombotic thrombocytopenic purpura.

TABLE 115-1 Drugs Reported as Definitely or Probably Causing Isolated Thrombocytopenia^a

Abciximab	Mirtazapine
Acetaminophen	Naproxen
Amiodarone	Oxaliplatin
Amlodipine	Penicillin
Ampicillin	Phenytoin
Carbamazepine	Piperacillin
Ceftriaxone	Quinidine
Cephalexin	Quinine
Ciprofloxacin	Ranitidine
Diazepam	Rosiglitazone
Eptifibatide	Roxifiban
Furosemide	Sulfisoxazole
Gold	Suramin
Haloperidol	Tirofiban
Heparin	Tranilast
Ibuprofen	Trimethoprim/sulfamethoxazole
Lorazepam	Vancomycin

^aBased on scoring requiring a compatible clinical picture and positive laboratory testing.

Source: Adapted from DM Arnold et al: J Thromb Hemost 11:169, 2013.

of naturally occurring antibodies that cross-react with the drug bound to the platelet.

Heparin-Induced Thrombocytopenia Drug-induced thrombocytopenia due to heparin differs from that seen with other drugs in two major ways. (1) The thrombocytopenia is not usually severe, with nadir counts rarely <20,000/ μ L. (2) Heparin-induced thrombocytopenia (HIT) is not associated with bleeding and, in fact, markedly increases the risk of thrombosis. The pathogenesis of HIT is complex. It results from antibody formation to a complex of the platelet-specific protein platelet factor 4 (PF4) and heparin or other glycosaminoglycans. The anti-heparin/PF4 antibody can activate platelets through the Fc γ RIIA receptor and also activate monocytes, endothelial cells, and coagulation proteins. Many patients exposed to heparin develop antibodies to heparin/PF4 but do not appear to have adverse consequences. A fraction of those who develop antibodies will develop HIT, and a portion of those (up to 50%) will develop thrombosis (HITT).

HIT can occur after exposure to low-molecular-weight heparin (LMWH) as well as unfractionated heparin (UFH), although it is more common with the latter. Most patients develop HIT after exposure to heparin for 5–14 days (Fig. 115-3). It occurs before 5 days in those who were exposed to heparin in the prior few weeks or months (<~100 days) and have circulating anti-heparin/PF4 antibodies. Rarely, thrombocytopenia and thrombosis begin several days after all heparin has been stopped (termed *delayed-onset HIT*), and more rarely, spontaneous HIT, or autoimmune HIT syndrome, occurs where there is no history of heparin exposure and termed *vaccine-induced immune thrombocytopenia and thrombosis* (VITT). A syndrome similar to spontaneous HIT has been described rarely post-COVID-19 vaccination mainly with the ChAdOx1-S/nCoV-19 vaccine. The “4T’s” have been recommended to be used in a diagnostic algorithm for HIT: thrombocytopenia, timing of platelet count drop, thrombosis and other sequelae such as localized skin reactions, and other causes of thrombocytopenia not evident. Application of the 4T scoring system is very useful in excluding a diagnosis of HIT but will result in overdiagnosis of HIT in situations where thrombocytopenia and thrombosis due to other etiologies are common, such as in the intensive care unit. Alternative scoring systems have been recommended, including for patients after cardiopulmonary bypass.

LABORATORY TESTING FOR HIT Because of the prevalence of antiheparin antibodies without clinical disease, testing should be done in individuals who are at intermediate or high risk based on clinical pretest assessment. HIT (anti-heparin/PF4) antibodies can be detected using two types of assays. The most widely available is an enzyme-linked immunoassay (ELISA) with PF4/polyanion complex as the antigen. Because many patients develop antibodies but do not develop clinical HIT, the test has a low specificity for the diagnosis of HIT. This is especially true in patients who have undergone surgery requiring cardiopulmonary bypass, where approximately 50% of patients develop these antibodies postoperatively. IgG-specific ELISAs increase specificity but may decrease sensitivity. The other assay is a platelet activation assay, most commonly the serotonin release assay, which measures the ability of the patient's serum to activate platelets in the presence of heparin in a concentration-dependent manner. This test has lower sensitivity

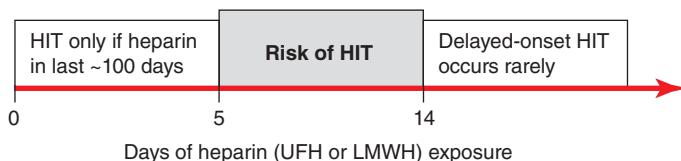


FIGURE 115-3 Time course of heparin-induced thrombocytopenia (HIT) development after heparin exposure. The timing of development after heparin exposure is a critical factor in determining the likelihood of HIT in a patient. HIT occurs early after heparin exposure in the presence of preexisting heparin/platelet factor 4 (PF4) antibodies, which disappear from circulation by ~100 days following a prior exposure. Rarely, HIT may occur later after heparin exposure (termed *delayed-onset HIT*). In this setting, heparin/PF4 antibody testing is usually markedly positive. HIT can occur after exposure to either unfractionated (UFH) or low-molecular-weight heparin (LMWH).

but higher specificity than the ELISA. However, HIT remains a clinical diagnosis.

TREATMENT

Heparin-Induced Thrombocytopenia

Early recognition is key in treatment of HIT, with prompt discontinuation of heparin and use of alternative anticoagulants if bleeding risk does not outweigh thrombotic risk. Thrombosis is a common complication of HIT, even after heparin discontinuation, and can occur in both the venous and arterial systems. In patients diagnosed with HIT, imaging studies to evaluate the patient for thrombosis (at least lower extremity duplex Doppler imaging) are recommended. Patients requiring anticoagulation should be switched from heparin to an alternative anticoagulant. The direct thrombin inhibitor (DTI) argatroban is effective in HITT. The DTI bivalirudin and the antithrombin-binding pentasaccharide fondaparinux are also effective but not approved by the U.S. Food and Drug Administration (FDA) for this indication. Direct oral anticoagulants (DOACs) are being used for treatment, although they are not FDA approved for this indication. Studies in small numbers of patients suggest their use in this setting may be a viable option. HIT antibodies cross-react with LMWH, and these drugs should not be used in the treatment of HIT.

Because of the high rate of thrombosis in patients with HIT, anticoagulation should be considered, even in the absence of thrombosis. In patients with thrombosis, anticoagulation is continued for 3–6 months, but in patients without thrombosis, the duration of anticoagulation is less well defined. An increased risk of thrombosis is present for at least 1 month after diagnosis; however, most thromboses occur early, and whether thrombosis occurs later if the patient is initially anticoagulated is unknown. Options include continuing anticoagulation until a few days after platelet recovery or for 1 month. Introduction of warfarin alone in the setting of HIT or HITT may precipitate thrombosis, particularly venous gangrene, presumably due to clotting activation and severely reduced levels of proteins C and S. Warfarin therapy, if started, should be overlapped with a DTI or fondaparinux and started after resolution of the thrombocytopenia and lessening of the prothrombotic state. Evidence for use of an oral direct Xa inhibitor in this setting is growing.

The rare VITT syndrome is characterized by high D-dimer levels and thrombosis in unusual sites like the cerebral venous sinuses. Fatal in about 20%, treatment is usually IgIV to block platelet activation through Fc receptors, the pathogenic effect of the anti-PF4-polyanion antibody.

Immune Thrombocytopenic Purpura Immune thrombocytopenic purpura (ITP; also termed *idiopathic thrombocytopenic purpura*) is an acquired disorder in which there is immune-mediated destruction of platelets and possibly inhibition of platelet release from the megakaryocyte. In children, it is usually an acute disease, most commonly following an infection, and with a self-limited course. In adults, it is a more chronic disease, although in some adults, spontaneous remission occurs, usually within months of diagnosis. ITP is termed *secondary* if it is associated with an underlying disorder; autoimmune disorders, particularly systemic lupus erythematosus (SLE), and infections, such as HIV and hepatitis C, are common causes. The association of ITP with *Helicobacter pylori* infection is unclear but appears to have a geographic distribution.

ITP is characterized by mucocutaneous bleeding and a low, often very low, platelet count, with an otherwise normal peripheral blood cells and smear. Patients usually present either with ecchymoses and petechiae or with thrombocytopenia incidentally found on a routine CBC. Mucocutaneous bleeding, such as oral mucosa, gastrointestinal, or heavy menstrual bleeding, may be present. Rarely, life-threatening, including central nervous system, bleeding can occur. Wet purpura (blood blisters in the mouth) and retinal hemorrhages may herald life-threatening bleeding.

LABORATORY TESTING IN ITP Laboratory testing for antibodies (serologic testing) is usually not helpful due to the low sensitivity and specificity of the current tests. Bone marrow examination can be reserved for those who have other signs or laboratory abnormalities not explained by ITP or in patients who do not respond to initial therapy. The peripheral blood smear may show large platelets, with otherwise normal morphology. Depending on the bleeding history, iron-deficiency anemia may be present.

Laboratory testing is performed to evaluate for secondary causes of ITP and should include testing for HIV infection and hepatitis C (and other infections if indicated). Serologic testing for SLE, serum protein electrophoresis, immunoglobulin levels to potentially detect hypogammaglobulinemia, selective testing for IgA deficiency or monoclonal gammopathies, and testing for *H. pylori* infection should be considered, depending on the clinical circumstance. If anemia is present, direct antiglobulin testing (Coombs' test) should be performed to rule out combined autoimmune hemolytic anemia with ITP (Evans' syndrome).

TREATMENT

Immune Thrombocytopenic Purpura

The treatment of ITP uses drugs that decrease reticuloendothelial uptake of the antibody-bound platelet, decrease antibody production, and/or increase platelet production. The diagnosis of ITP does not necessarily mean that treatment must be instituted. Patients with platelet counts >30,000/ μ L appear not to have increased mortality related to the thrombocytopenia.

Initial treatment in patients without significant bleeding symptoms, severe thrombocytopenia (<5000/ μ L), or signs of impending bleeding (e.g., retinal hemorrhage or large oral mucosal hemorrhages) can be instituted as an outpatient using single agents. Traditionally, this has been prednisone at 1 mg/kg or a 4-day course of dexamethasone, 40 mg/d, although Rh₀(D) immune globulin therapy (WinRho SDF), at 50–75 μ g/kg, is also being used in this setting. Rh₀(D) immune globulin must be used only in Rh-positive patients because the mechanism of action is production of limited hemolysis, with antibody-coated cells "saturating" the Fc receptors, inhibiting Fc receptor function. Monitoring patients for 8 h after infusion is now advised by the FDA because of the rare complication of severe intravascular hemolysis. Intravenous gamma globulin (IVIgG), which is pooled, primarily IgG antibodies, also blocks the Fc receptor system, but appears to work primarily through different mechanism(s). IVIgG has more efficacy than anti-Rh₀(D) in post-splenectomy patients. IVIgG is dosed at 1–2 g/kg total, given over 1–5 days. Side effects are usually related to the volume of infusion and infrequently include aseptic meningitis and renal failure. All immunoglobulin preparations are derived from human plasma and undergo treatment for viral inactivation.

For patients with severe ITP and/or symptoms of bleeding, hospital admission is required, and combined-modality therapy is given using high-dose glucocorticoids with IVIgG or anti-Rh₀(D) therapy and, as needed, additional immunosuppressive agents. Rituximab, an anti-CD20 (B cell) antibody, has shown efficacy in the treatment of refractory ITP, although long-lasting remission only occurs in approximately 30% of patients.

TPO receptor agonists, one administered subcutaneously (romiprolast) and another orally (eltrombopag), are effective in raising platelet counts in patients with ITP and are recommended for patients who relapse or who are unresponsive to at least one other therapy.

Other immunosuppressive drugs have also been tested. The combination of glucocorticoids with mycophenolate mofetil (500 mg PO bid, increasing to 1000 mg PO bid as tolerated) appears to be more effective than glucocorticoids alone.

Splenectomy has been used for treatment of patients who relapse after glucocorticoids are tapered and remains a treatment option. However, with the recognition that ITP will resolve spontaneously in some adult patients, observation, if the platelet count is high enough, or intermittent treatment with anti-Rh₀(D) or IVIgG, or

initiation of treatment with a TPO receptor agonist may be a reasonable approach to see if the ITP will resolve, prior to splenectomy or other therapies. Vaccination against encapsulated organisms (especially pneumococcus, but also meningococcus and *Haemophilus influenzae*, depending on patient age and potential exposure) is recommended before splenectomy. Accessory spleens are a very rare cause of relapse.

Inherited Thrombocytopenia Thrombocytopenia is rarely inherited, either as an isolated finding or as part of a syndrome, and may be inherited in an autosomal dominant, autosomal recessive, or X-linked pattern. Many forms of autosomal dominant macrothrombocytopenia are now known to be associated with variants in the non-muscle myosin heavy chain *MYH9* gene. Interestingly, these include the May-Hegglin anomaly, and Sebastian, Epstein's, and Fechtner syndromes, all of which have distinct distinguishing features. A common feature of these disorders is large platelets (Fig. 115-1C). Autosomal recessive disorders include congenital amegakaryocytic thrombocytopenia, thrombocytopenia with absent radii, and Bernard-Soulier syndrome. The latter is primarily a functional platelet disorder due to absence of Gp Ib-IX-V, the VWF adhesion receptor. X-linked disorders include Wiskott-Aldrich syndrome and a dyshematopoietic syndrome resulting from a mutation in *GATA-1*, an important transcriptional regulator of hematopoiesis.

■ THROMBOTIC THROMBOCYTOPENIC PURPURA AND HEMOLYTIC-UREMIC SYNDROME

Thrombotic thrombocytopenic microangiopathies are a group of disorders characterized by microangiopathic hemolytic anemia (MAHA) defined by thrombocytopenia and fragmented RBCs (Fig. 115-1D) on peripheral blood smear, laboratory evidence of hemolysis (elevated lactate dehydrogenase [LDH] and unconjugated bilirubin and decreased haptoglobin), and microvascular thrombosis. They include thrombotic thrombocytopenic purpura (TTP) and hemolytic-uremic syndrome (HUS), as well as syndromes complicating bone marrow transplantation, certain medications and infections, pregnancy, and vasculitis. In DIC, although thrombocytopenia and microangiopathy are seen, a coagulopathy predominates, with consumption of clotting factors and fibrinogen resulting in an elevated prothrombin time (PT) and often activated partial thromboplastin time (aPTT). The PT and aPTT are characteristically normal in TTP or HUS.

Thrombotic Thrombocytopenic Purpura TTP was first described in 1924 by Eli Moschcowitz and characterized by a pentad of findings that include microangiopathic hemolytic anemia, thrombocytopenia, renal failure, neurologic findings, and fever. The full-blown syndrome is less commonly seen now, probably due to earlier diagnosis. The introduction of treatment with plasma exchange markedly improved the prognosis in patients, with a decrease in mortality from 85–100% to 10–30%.

The pathogenesis of inherited (Upshaw-Schulman syndrome) and idiopathic TTP (ITTP) is related to a deficiency of, or antibodies to, the metalloprotease ADAMTS13, which cleaves VWF. VWF is normally secreted as ultra-large multimers, which are then cleaved by ADAMTS13. The persistence of ultra-large VWF molecules is thought to contribute to pathogenic platelet adhesion and aggregation (Fig. 115-4). This defect alone, however, is not sufficient to result in TTP because individuals with a congenital absence of ADAMTS13 develop TTP only episodically, including during first pregnancy. The level of ADAMTS13 activity, as well as antibodies to ADAMTS13, can be detected by laboratory assays, which play a critical role in the differential diagnosis of MAHA. ADAMTS13 activity levels of <10% are diagnostic of TTP.

Idiopathic TTP appears to be more common in women than in men. No geographic or racial distribution has been defined. TTP is more common in patients with HIV infection and in pregnant women. Medication-related MAHA may be secondary to antibody formation (ticlopidine and possibly clopidogrel) or direct endothelial toxicity (cyclosporine, mitomycin C, tacrolimus, quinine), although this is not always so clear, and fear of withholding treatment, as well as lack of

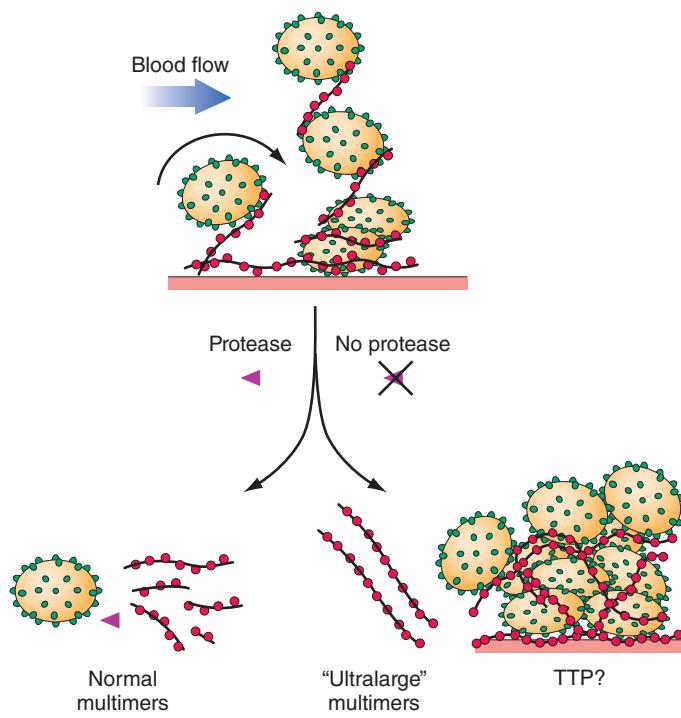
VWF and Platelet Adhesion

FIGURE 115-4 Pathogenesis of thrombotic thrombocytopenic purpura (TTP). Normally the ultra-high-molecular-weight multimers of von Willebrand factor (VWF) produced by the endothelial cells are processed into smaller multimers by a plasma metalloproteinase called ADAMTS13. In TTP, the activity of the protease is inhibited, and the ultra-high-molecular-weight multimers of VWF initiate platelet aggregation and thrombosis.

other treatment alternatives, may result in initial application of plasma exchange. However, withdrawal, or reduction in dose, of endothelial toxic agents usually decreases the microangiopathy.

TREATMENT**Thrombotic Thrombocytopenic Purpura**

TTP is a devastating disease if not diagnosed and treated promptly. In patients presenting with new thrombocytopenia, with or without evidence of renal insufficiency and other elements of classic TTP, laboratory data (PT, aPTT, CBC with platelet count and peripheral smear, ADAMTS13 activity, LDH, bilirubin, haptoglobin, direct antiglobulin assay) should be obtained to rule out DIC and to evaluate for evidence of MAHA.

Therapeutic plasma exchange (TPE) remains the mainstay of treatment of TTP. TPE is continued until the platelet count is normal and signs of hemolysis are resolved for at least 2 days. Although never evaluated in clinical trials, the use of glucocorticoids seems a reasonable approach but should only be used as an adjunct to plasma exchange. The addition of rituximab to initial therapy decreases duration of TPE and relapses. Caplacizumab, an anti-VWF nanobody, decreases mortality and burden of care when used in patients with ADAMTS13 <10% or with high clinical probability of disease. Guidelines from the International Society of Thrombosis and Hemostasis recommend starting caplacizumab and rituximab only in individuals with diagnostic ADAMTS13 levels (usually <10%) and, additionally for rituximab, in patients with evidence of an inhibitor, given potential side effects and costs.

Patients with persistently low ADAMTS13 have a greater risk of ongoing sequelae including stroke. There is a significant relapse rate; in patients treated with TPE, 25–45% of patients relapse within 30 days of initial “remission,” and 12–40% of patients have late relapses. Relapses are more frequent in patients with severe ADAMTS13 deficiency at presentation. Treatment of patients with

TTP relapses should be initiated before confirmatory laboratory assays are available.

Hemolytic-Uremic Syndrome HUS is a syndrome characterized by acute renal failure, microangiopathic hemolytic anemia, and thrombocytopenia. It is seen preceded by an episode of diarrhea, often hemorrhagic in nature, predominantly in children. *Escherichia coli* O157:H7 is the most frequent, although not only, etiologic serotype. HUS not associated with diarrhea is more heterogeneous in presentation and course. Atypical HUS (aHUS) is usually due to genetic defects in complement genes or antibodies directed against complementary regulatory proteins that result in chronic complement activation. Laboratory testing for DNA variants in complement regulatory genes is available, although assigning pathogenicity to variants remains challenging. Currently, a commercially available functional assay is not available that is diagnostic of the disease.

TREATMENT**Hemolytic-Uremic Syndrome**

Treatment of HUS is primarily supportive. In HUS associated with diarrhea, many (~40%) children require at least some period of support with dialysis; however, the overall mortality is <5%. In HUS not associated with diarrhea, the mortality is higher, approximately 26%. Plasma infusion or plasma exchange has not been shown to alter the overall course in HUS or aHUS, except in patients with antibodies to factor H. ADAMTS13 levels are generally reported to be normal in HUS, although occasionally they have been reported to be decreased. In patients with aHUS, eculizumab, a humanized monoclonal antibody against C5 that blocks terminal complement, has efficacy in resolution of aHUS and improving or preserving renal function. Patients with aHUS may initially be treated with plasma exchange, until the ADAMTS13 level is returned and the diagnosis is more clear, since aHUS remains a diagnosis of exclusion. However, plasma exchange has not been shown to affect clinical outcomes in aHUS.

THROMBOCYTOSIS

Thrombocytosis is almost always due to (1) iron deficiency; (2) inflammation, cancer, or infection (reactive thrombocytosis); or (3) an underlying myeloproliferative process (essential thrombocythemia or polycythemia vera) (Chap. 103) or, rarely, the 5q- myelodysplastic process (Chap. 102). Patients presenting with an elevated platelet count should be evaluated for underlying inflammation and malignancy, and iron deficiency should be ruled out. Thrombocytosis in response to acute or chronic inflammation has not been clearly associated with an increased thrombotic risk. In fact, patients with markedly elevated platelet counts (>1.5 million), usually seen in the setting of a myeloproliferative disorder, have an increased risk of bleeding. This appears to be due, at least in part, to acquired von Willebrand disease (VWD) due to platelet-VWF binding and removal from the circulation.

QUALITATIVE DISORDERS OF PLATELET FUNCTION

Inherited Disorders of Platelet Function Inherited platelet function disorders are thought to be relatively rare, although the prevalence of mild disorders of platelet function is unclear, in part because our testing for such disorders is suboptimal. Rare qualitative disorders include the autosomal recessive disorders Glanzmann's thrombasthenia (absence of the platelet Gp IIb/IIIa receptor) and Bernard-Soulier syndrome (absence of the platelet Gp Ib-IX-V receptor). Both are inherited in an autosomal recessive fashion and present with bleeding symptoms in childhood.

Platelet storage pool disorder (SPD) is the classic autosomal dominant qualitative platelet disorder. This results from abnormalities of platelet granule formation. It is also seen as a part of inherited disorders of granule formation, such as Hermansky-Pudlak syndrome. Bleeding

symptoms in SPD are variable but often are mild. The most common inherited disorders of platelet function prevent normal secretion of granule content and are termed *secretion defects*. An increasing number of genetic variants are being found in patients with these disorders, although assigning pathogenicity remains challenging.

TREATMENT

Inherited Disorders of Platelet Dysfunction

Bleeding symptoms or prevention of bleeding in patients with severe platelet dysfunction frequently requires platelet transfusion. Care must be taken to limit the risk of alloimmunization by limiting exposure and using HLA-matched single donor platelets for transfusion when needed. rFVIIa is FDA approved in Glanzmann's thrombasthenia and Bernard Soulier syndrome where use can avoid platelet alloimmunization and anti-receptor antibody formation. Platelet disorders associated with milder bleeding symptoms frequently respond to desmopressin (1-deamino-8-D-arginine vasopressin [DDAVP]). DDAVP increases plasma VWF and factor VIII levels; it may also have a direct effect on platelet function. Particularly for mucosal bleeding symptoms, antifibrinolytic therapy (tranexamic acid or ϵ -aminocaproic acid) is used alone or in conjunction with DDAVP or platelet therapy.

Acquired Disorders of Platelet Function Acquired platelet dysfunction is common, usually due to medications, either intentionally as with antiplatelet therapy or unintentionally as with high-dose penicillins. Acquired platelet dysfunction occurs in uremia. This is likely multifactorial, but the resultant effect is defective adhesion and activation. The platelet defect is improved most by dialysis but may also be improved by increasing the hematocrit to 27–32%, giving DDAVP (0.3 $\mu\text{g}/\text{kg}$), or use of conjugated estrogens. Platelet dysfunction also occurs with cardiopulmonary bypass due to the effect of the artificial circuit on platelets, and bleeding symptoms respond to platelet transfusion. Platelet dysfunction seen with underlying hematologic disorders can result from nonspecific interference by circulating paraproteins or intrinsic platelet defects in myeloproliferative and myelodysplastic syndromes.

VON WILLEBRAND DISEASE

VWD is the most common inherited bleeding disorder, with prevalence of symptomatic disease of 1 in 1000 to 1 in 10,000 individuals. VWF serves two roles: (1) as the major adhesion molecule that tethers the platelet to the exposed subendothelium; and (2) as the binding protein for factor VIII (FVIII), resulting in significant prolongation of the FVIII half-life in circulation. The platelet-adhesive function of VWF is critically dependent on the presence of large VWF multimers, whereas FVIII binding is not. Most of the symptoms of VWD are “platelet-like” except in more severe VWD when the FVIII is low enough to produce symptoms similar to those found in FVIII deficiency (hemophilia A).

VWD has been classified into three major types, with four subtypes of type 2 (Table 115-2). By far, the most common type of VWD is type 1 disease, with a parallel decrease in VWF protein, VWF function, and FVIII levels, accounting for at least 80% of cases. In type 1 VWD, patients have predominantly mucosal bleeding symptoms, although postoperative bleeding can also be seen. Bleeding symptoms are uncommon in infancy and usually manifest later in childhood with excessive bruising and epistaxis. Because these symptoms occur commonly in childhood, the clinician should particularly note bruising at sites unlikely to be traumatized and/or prolonged epistaxis requiring medical attention. Heavy menstrual bleeding is a common manifestation of VWD. Menstrual bleeding resulting in anemia should warrant an evaluation for VWD and, if negative, functional platelet disorders. Type 1 VWD may first manifest with dental extractions, particularly wisdom tooth extraction, or tonsillectomy.

Not all patients with low VWF levels have bleeding symptoms. Whether patients bleed or not will depend on the overall hemostatic balance they have inherited, along with environmental influences and the type of hemostatic challenges they experience. Although the

TABLE 115-2 Laboratory Diagnosis of von Willebrand Disease (VWD)

TYPE	aPTT	VWF ANTIGEN	VWF ACTIVITY	FVIII ACTIVITY	MULTIMER
1	NI or ↑	↓	↓	↓	Normal distribution, decreased in quantity
2A	NI or ↑	↓	↓↓	↓	Loss of high- and intermediate-MW multimers
2B ^a	NI or ↑	↓	↓↓	↓	Loss of high-MW multimers
2M	NI or ↑	↓	↓↓	↓	Normal distribution, decreased in quantity
2N	↑↑	NI or ↓ ^b	NI or ↓ ^b	↓↓	Normal distribution
3	↑↑	↓↓	↓↓	↓↓	Absent

^aUsually also decreased platelet count. ^bFor type 2N, in the homozygous state, factor VIII is very low; in the heterozygous state, it is only seen in conjunction with type 1 VWD.

Abbreviations: aPTT, activated partial thromboplastin time; F, factor; MW, molecular weight; NI, normal; VWF, von Willebrand factor.

inheritance of VWD is autosomal, many factors modulate both VWF levels and bleeding symptoms. These have not all been defined, but include blood type, thyroid hormone status, race, stress, exercise, hormonal (both endogenous and exogenous) influences, and modulators of VWF clearance. Patients with type O blood have VWF protein levels of approximately one-half those of patients with AB blood type, and in fact, the normal range for patients with type O blood overlaps that which has been considered diagnostic for VWD. Patients with mildly decreased VWF levels should be diagnosed with VWD only in the setting of bleeding symptoms and/or a family history of VWD.

Patients with type 2 VWD have functional defects; thus, the VWF antigen measurement is significantly higher than the test of function. For types 2A, 2B, and 2M VWD, platelet-binding and/or collagen-binding VWF activity is decreased. In type 2A VWD, the impaired function is due either to increased susceptibility to cleavage by ADAMTS13, resulting in loss of intermediate- and high-molecular-weight multimers, or to decreased production of these multimers by the cell. Type 2B VWD results from gain-of-function DNA variants that result in increased ADAMTS13 cleavage and binding of VWF to platelets in circulation, with subsequent clearance of this complex by the reticuloendothelial system. The resulting VWF in the patients' plasma lacks the highest molecular-weight multimers, and the platelet count is usually modestly reduced. Type 2M occurs as a consequence of a group of DNA variants that cause dysfunction but do not affect multimer structure.

Type 2N VWD is due to variants in the VWF gene that affect binding of FVIII. As FVIII is stabilized by binding to VWF, the FVIII in patients with type 2N VWD has a very short half-life, and the FVIII level is markedly decreased. This is sometimes termed *autosomal hemophilia*. Type 3 VWD, or severe VWD, describes patients with virtually no VWF protein and usually FVIII levels <10%. Patients experience mucosal and joint bleeding, surgery-related bleeding, and other bleeding symptoms. Some patients with type 3 VWD, particularly those with large VWF gene deletions, are at risk of developing antibodies to infused VWF.

Acquired VWD or von Willebrand syndrome is most commonly seen in patients with underlying lymphoproliferative disorders, including monoclonal gammopathies of undetermined significance (MGUS), multiple myeloma, and Waldenström's macroglobulinemia. It is seen most commonly in the setting of MGUS and should be suspected in patients, particularly elderly patients, with a new onset of severe mucosal bleeding symptoms. Laboratory evidence of acquired VWD is found in some patients with cardiac valvular disease. Heyde's syndrome (aortic stenosis with gastrointestinal bleeding) is attributed to the presence of angiogenesis of the gastrointestinal tract in patients with aortic stenosis. The shear stress on blood passing through the stenotic aortic valve appears to unfold VWF, making it susceptible to proteolysis. Consequently, large multimer forms are lost, leading to an acquired type 2 VWD, but return when the stenotic valve is replaced.

TREATMENT

Von Willebrand Disease

The mainstay of treatment for type 1 VWD is DDAVP (desmopressin), which results in release of VWF and FVIII from endothelial stores. DDAVP can be given intravenously, by high-concentration intranasal spray (1.5 mg/mL), or when a concentrated form is available, by subcutaneous injection. The peak activity when given intravenously is approximately 30 min, whereas it is 2 h when given intranasally. The usual dose is 0.3 µg/kg intravenously or two squirts (one in each nostril) for patients >50 kg (one squirt for those <50 kg). It is recommended that patients with VWD be tested with DDAVP to assess their response before using it. In patients who respond well (increase in laboratory values of two- to fourfold), it can be used for procedures with minor to moderate risk of bleeding. Depending on the procedure, additional doses may be needed; it is usually given every 12–24 h. Less frequent dosing may result in less tachyphylaxis, which occurs when synthesis cannot compensate for the released stores. The major side effect of DDAVP is hyponatremia due to decreased free water clearance. This occurs most commonly in the very young and the very old, but fluid restriction should be advised for all patients for the 24 h following each dose.

Some patients with types 2A VWD respond to DDAVP such that it can be used for minor procedures. For the other subtypes, for type 3 disease, and for major procedures requiring longer periods of normal hemostasis, VWF replacement can be given. Virally inactivated VWF-plasma-derived and recombinant factor concentrates are safer than cryoprecipitate as the replacement product.

Antifibrinolytic therapy using either tranexamic acid (TXA) or ϵ -aminocaproic acid is an important therapy, either alone or in an adjunctive capacity, particularly for the prevention or treatment of mucosal bleeding. These agents are particularly useful in treatment of heavy menstrual bleeding (TXA 1300 mg every 8 h) and postpartum hemorrhage, as prophylaxis for dental procedures, and with DDAVP or factor concentrate for dental extractions, tonsillectomies, and prostate procedures. Antifibrinolytic agents are contraindicated in the setting of upper urinary tract bleeding due to the risk of ureteral obstruction.

DISORDERS OF THE VESSEL WALL

The vessel wall is an integral part of hemostasis, and separation of a fluid phase is artificial, particularly in disorders such as TTP or HIT that clearly involve the endothelium as well. Inflammation localized to the vessel wall, such as vasculitis, and inherited connective tissue disorders are abnormalities inherent to the vessel wall.

Metabolic and Inflammatory Disorders Acute febrile illnesses may result in vascular damage. This can result from immune complexes containing viral antigens or the viruses themselves. Certain pathogens, such as the rickettsiae causing Rocky Mountain spotted fever, replicate in endothelial cells and damage them. SARS-CoV-2 also infects endothelial cells, resulting in activation and damage contributing to COVID-19 pathogenicity. Vascular purpura may occur in patients with polyclonal gammopathies but more commonly occurs in those with monoclonal gammopathies, including Waldenström's macroglobulinemia, multiple myeloma, and cryoglobulinemia. Patients with mixed cryoglobulinemia develop a more extensive maculopapular rash due to immune complex-mediated damage to the vessel wall.

Patients with scurvy (vitamin C deficiency) develop painful episodes of perifollicular skin bleeding as well as more systemic bleeding symptoms. Vitamin C is needed to synthesize hydroxyproline, an essential constituent of collagen. Patients with Cushing's syndrome or on chronic glucocorticoid therapy develop skin bleeding and easy bruising due to atrophy of supporting connective tissue. A similar phenomenon is seen with aging, where following minor trauma, blood spreads superficially under the epidermis. This has been termed *senile purpura*. It is most common on skin that has been previously damaged by sun exposure.

Henocho-Schönlein, or anaphylactoid, purpura is a distinct, self-limited type of vasculitis that occurs in children and young adults. Patients have an acute inflammatory reaction with IgA and complement components in capillaries, mesangial tissues, and small arterioles leading to increased vascular permeability and localized hemorrhage. The syndrome is often preceded by an upper respiratory infection, commonly with streptococcal pharyngitis, or is triggered by drug or food allergies. Patients develop a purpuric rash on the extensor surfaces of the arms and legs, usually accompanied by polyarthralgias or arthritis, abdominal pain, and hematuria from focal glomerulonephritis. All coagulation tests are normal, but renal impairment may occur. Glucocorticoids can provide symptomatic relief but do not alter the course of the illness.

Inherited Disorders of the Vessel Wall Patients with inherited disorders of the connective tissue matrix, such as Marfan's syndrome, Ehlers-Danlos syndrome, and pseudoxanthoma elasticum, frequently report easy bruising. Inherited vascular abnormalities can result in increased bleeding. This is notably seen in hereditary hemorrhagic telangiectasia (HHT, or Osler-Weber-Rendu disease), a disorder where abnormal telangiectatic capillaries result in frequent bleeding episodes, primarily from the nose and gastrointestinal tract. Arteriovenous malformation (AVM) in the lung, brain, and liver may also occur in HHT. The telangiectasia can often be visualized on the oral and nasal mucosa. Signs and symptoms develop over time. Epistaxis begins, on average, at the age of 12 and occurs in >95% of affected individuals by middle age. Approximately 25% have gastrointestinal bleeding usually beginning after the age of 50. HHT is caused by pathogenic DNA variants in number of genes involved in the TGF β /BMP signaling cascade.

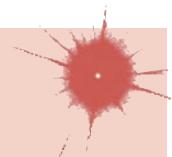
FURTHER READING

- BOENDER J et al: A diagnostic approach to mild bleeding disorders. *J Thromb Haemost* 14:1507, 2016.
- GREINACHER A et al: Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. *N Engl J Med* 384:2092, 2021.
- GRESELE P et al: Diagnosis of inherited platelet function disorders: Guidance from the SSC of the ISTH. *J Thromb Haemost* 13:314, 2015.
- HOGAN M, BERGER JS: Heparin-induced thrombocytopenia. Review of incidence, diagnosis and management. *Vasc Med* 25:160, 2020.
- HUNT BJ: Bleeding and coagulopathies in critical care. *N Engl J Med* 370:847, 2014.
- JAMES PD et al: ASH ISTH NHF WFH 2021 guidelines on the diagnosis of von Willebrand disease. *Blood Adv* 5:280, 2021.
- JOKIRANTA TS: HUS and atypical HUS. *Blood* 129:2847, 2017.
- LEEBECK FWG, EIKENBOOM JCJ: von Willebrand's disease. *N Engl J Med* 375:2067, 2016.
- SKEITH L et al: A practical approach to evaluating postoperative thrombocytopenia. *Blood Adv* 4:776, 2020.
- VAYNE C et al: Pathophysiology and diagnosis of drug-induced immune thrombocytopenia. *J Clin Med* 9:2212, 2020.
- ZHENG XL et al: ISTH guidelines for the diagnosis of thrombotic thrombocytopenic purpura. *J Thromb Hemost* 18:2486, 2020.

116

Coagulation Disorders

Jean M. Connors



Deficiencies of coagulation factors have been recognized for centuries. Patients with genetic deficiencies of plasma coagulation factors exhibit lifelong recurrent bleeding episodes into joints, muscles, and closed spaces, either spontaneously or following an injury. The most common inherited factor deficiencies are the hemophilias, X-linked diseases caused by deficiency of factor (F) VIII (hemophilia A) or FIX

TABLE 116-1 Genetic and Laboratory Characteristics of Inherited Coagulation Disorders

CLOTTING FACTOR DEFICIENCY	INHERITANCE	PREVALENCE IN GENERAL POPULATION	LABORATORY ABNORMALITY ^a			MINIMUM HEMOSTATIC LEVELS	TREATMENT	PLASMA HALF-LIFE
			aPTT	PT	TT			
Fibrinogen	AR	1 in 1,000,000	+	+	+	100 mg/dL	Cryoprecipitate	2–4 d
Prothrombin	AR	1 in 2,000,000	+	+	-	20%–30%	FFP/PCC	3–4 d
Factor V	AR	1 in 1,000,000	+/-	+/-	-	15%–20%	FFP ^c	36 h
Factor VII	AR	1 in 500,000	-	+	-	15%–20%	FFP/PCC	4–6 h
Factor VIII	X-linked	1 in 5000	+	-	-	30%	FVIII concentrates	8–12 h
Factor IX	X-linked	1 in 30,000	+	-	-	30%	FIX concentrates	18–24 h
Factor X	AR	1 in 1,000,000	+/-	+/-	-	15%–20%	FFP/PCC	40–60 h
Factor XI	AR	1 in 1,000,000	+	-	-	15%–20%	FFP	40–70 h
Factor XII	AR	ND	+	-	-	b	b	60 h
HK	AR	ND	+	-	-	b	b	150 h
Prekallikrein	AR	ND	+	-	-	b	b	35 h
Factor XIII	AR	1 in 2,000,000	-	-	+/-	2%–5%	Cryoprecipitate/FXIII concentrates	11–14 d

^aValues within normal range (-) or prolonged (+). ^bNo risk for bleeding; treatment is not indicated. ^cSince platelets contain FV, platelet transfusion can be used as therapy.

Abbreviations: aPTT, activated partial thromboplastin time; AR, autosomal recessive; FFP, fresh-frozen plasma; HK, high-molecular-weight kininogen; ND, not determined; PCC, prothrombin complex concentrates; PT, prothrombin time; TT, thrombin time.

(hemophilia B). Rare congenital bleeding disorders due to deficiencies of other factors, including FII (prothrombin), FV, FVII, FX, FXI, FXIII, and fibrinogen, are commonly inherited in an autosomal recessive manner (Table 116-1). Disease phenotype often correlates with the level of factor activity. While patients can have a congenital deficiency of FXII accompanied by a significant prolongation in the activated partial thromboplastin time (aPTT), FXII deficiency is not accompanied by a bleeding phenotype, likely due to redundant paths to activation of the intrinsic pathway of the coagulation cascade, including direct activation of FXI by thrombin generated through the extrinsic pathway (Fig. 116-1). Advances in characterization of the molecular basis of clotting factor deficiencies have contributed to better understanding of the disease phenotypes allowing the development of more targeted therapeutic approaches, including the use of small molecules, recombinant proteins, or cell- and gene-based therapies.

The two most commonly used tests of hemostasis, the prothrombin time (PT) and the aPTT, were designed to perform the first screen for clotting factor deficiency (Fig. 116-1). An isolated prolonged PT suggests FVII deficiency, whereas a prolonged aPTT indicates an intrinsic pathway factor deficiency, most commonly hemophilia A or B (FVIII or FIX, respectively) or FXI deficiency (Fig. 116-1). The prolongation of both PT and aPTT suggests a deficiency of FV, FX, FII, or fibrinogen abnormalities. A mixing study, in which the addition of normal pooled plasma to the patient's plasma, will correct a prolonged aPTT or PT due to a factor deficiency, and is the next step in determining if there is a coagulation factor deficiency. If the clotting time does not correct, it suggests the presence of an inhibitor, an antibody to a specific factor; however, a mixing study will also detect the presence of anticoagulants. Many labs have testing methods for detecting inhibitors that neutralize anticoagulants. If the mixing study corrects with normal plasma, individual factor activity assays are performed to determine which factor is deficient.

Acquired deficiencies of plasma coagulation factors are more frequent than congenital disorders; the most common disorders include hemorrhagic diathesis of liver disease, disseminated intravascular coagulation (DIC), and vitamin K deficiency. In these disorders, blood coagulation

is hampered by the deficiency of more than one clotting factor, and the bleeding episodes are the result of perturbation of both primary (e.g., platelet and vessel wall interactions) and secondary (coagulation) hemostasis.

The development of alloantibodies to coagulation plasma proteins, clinically termed *inhibitors*, is a relatively rare disease that often affects hemophilia A or B and FXI-deficient patients on repetitive exposure to the missing protein to control bleeding episodes. Inhibitory autoantibodies also occur among subjects without genetic deficiency of clotting factors and although rare can be seen in the postpartum setting, as a manifestation of underlying autoimmune or neoplastic disease, or idiopathically. Rare cases of acquired inhibitors to thrombin or FV have been reported in patients receiving topical bovine thrombin preparation as a local hemostatic agent in complex surgeries. The results of a mixing study that does not correct with the addition of normal plasma indicate the presence of an inhibitor, requiring additional tests to identify the specificity of the inhibitor and measure its titer. Inhibitor

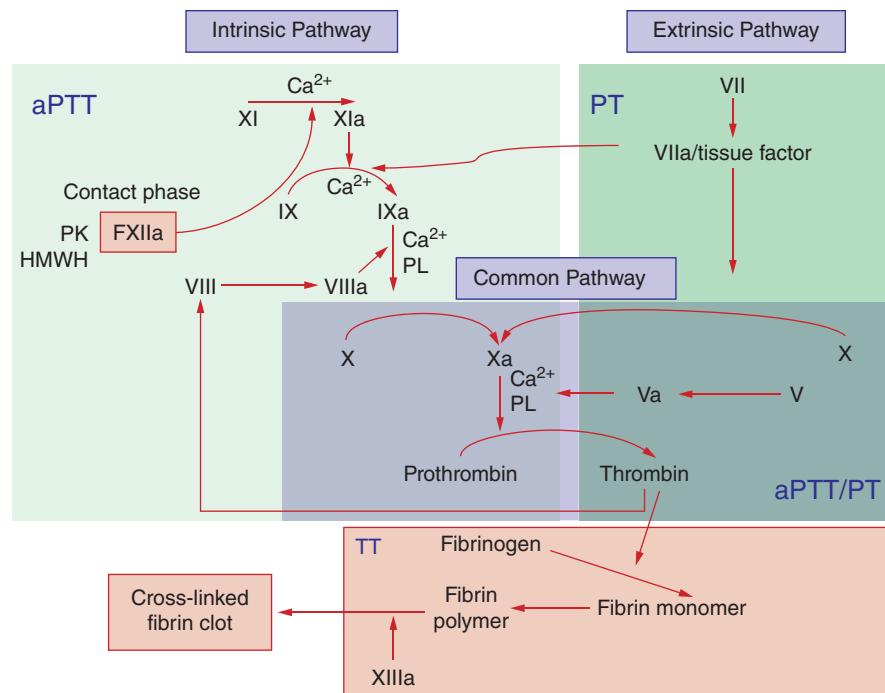


FIGURE 116-1 Coagulation cascade and laboratory assessment of clotting factor deficiency by activated partial thromboplastin time (aPTT), prothrombin time (PT), thrombin time (TT), and phospholipid (PL).

detection in patients with hemophilia is of particular importance, with yearly screening performed at most hemophilia treatment centers.

The treatment of coagulation factor deficiencies in the setting of bleeding requires replacement of the deficient protein(s) using recombinant or purified plasma-derived products or fresh-frozen plasma (FFP). Prothrombin complex concentrates (PCCs) are intermediate purity plasma-derived factor concentrates initially used as sources of FVIII or FIX for hemophilia patients, but as they contain the vitamin K-dependent factors, are also used for warfarin reversal. Three-factor PCC (3F-PCC) is less frequently used now for warfarin reversal because these preparations contain low levels of FVII, requiring FFP as a source of FVII. Four-factor PCC (4F-PCC), especially the one used in the United States, contains FII, FIX, FX, higher levels of FVII than 3F-PCC, and protein S and protein C.

HEMOPHILIA A AND B

■ PATHOGENESIS AND CLINICAL MANIFESTATIONS

Hemophilia is an X-linked recessive hemorrhagic disease due to mutations in the *F8* gene (hemophilia A or classic hemophilia) or *F9* gene (hemophilia B). The disease affects 1 in 10,000 males worldwide, in all ethnic groups; hemophilia A represents 80% of all cases. The large size of the *F8* gene makes it more susceptible to mutation events than the smaller *F9* gene. Male subjects are clinically affected; women, who carry a single mutated gene, are generally asymptomatic. Family history of the disease is absent in ~30% of cases, and in these cases 80% of the mothers are carriers of the de novo mutated allele. More than 500 different mutations have been identified in the *F8* or *F9*. One of the most common hemophilia A mutations results from an inversion of the intron 22 sequence, and it is present in 40% of cases of severe hemophilia A. Advances in molecular diagnosis now permit precise identification of mutations, allowing accurate diagnosis of women carriers of the hemophilia gene in affected families.

Clinically, hemophilia A and hemophilia B are indistinguishable. The disease phenotype correlates with the activity of FVIII or FIX and can be classified as severe (<1%), moderate (1–5%), or mild (6–30%). In the severe and moderate forms, the disease is characterized by bleeding into the joints (hemarthrosis), soft tissues, and muscles after minor trauma or even spontaneously. Patients with mild disease experience infrequent bleeding that is usually secondary to trauma. Among those with residual FVIII or FIX activity >25% of normal, the disease is discovered only in the event of bleeding after major trauma or during routine preoperative laboratory tests, usually with an isolated prolongation of the aPTT that requires further investigation with a mixing study. Factor VIII has a short circulating half-life of 25–30 min that is extended to roughly 12 h when complexed with its carrier protein von Willebrand factor (VWF). In patients without a known history of hemophilia, a diagnosis of von Willebrand disease (VWD) needs to be excluded in patients with a prolonged aPTT and low FVIII activity. Early in life, bleeding may present after circumcision or rarely as intracranial hemorrhages. The disease is more evident when children begin to walk or crawl. In the severe form, the most common bleeding manifestations are recurrent hemarthroses, affecting primarily the knees, elbows, ankles, shoulders, and hips. Acute hemarthroses are painful, and clinical signs are local swelling and erythema. To avoid pain, the patient may adopt a fixed position, which leads eventually to muscle contractures. Very young children unable to communicate verbally show irritability and a lack of movement of the affected joint. Chronic hemarthroses are debilitating with synovial thickening and synovitis in response to the intraarticular blood. After a joint has been damaged, recurrent bleeding episodes result in the clinically recognized “target joint,” which then establishes a vicious cycle of bleeding, resulting in progressive joint deformity that in critical cases requires surgery as the only therapeutic option. Hematomas into the muscle of distal parts of the limbs may lead to external compression of arteries, veins, or nerves that can result in compartment syndrome.

Bleeding into the oropharyngeal spaces, central nervous system (CNS), or retroperitoneum is life-threatening and requires immediate therapy. Retroperitoneal hemorrhages can accumulate large quantities

of blood with formation of masses with calcification and inflammatory tissue reaction (pseudotumor syndrome) and also result in damage to the femoral nerve. Pseudotumors can also form in bones, especially long bones of the lower limbs. Hematuria is frequent among hemophilia patients, even in the absence of genitourinary pathology. It is often self-limited and may not require specific therapy.

TREATMENT

Hemophilia

Without treatment, severe hemophilia may limit life expectancy. Advances in the blood fractionation industry during World War II resulted in the realization that plasma could be used to treat hemophilia, but the volumes required to achieve even modest elevation of circulating factor levels limit the utility of plasma infusion as an approach to disease management. The discovery in the 1960s that the cryoprecipitate fraction of plasma was enriched for FVIII, and the eventual purification of FVIII and FIX from plasma, led to the introduction of home infusion therapy with factor concentrates in the 1970s. The availability of factor concentrates resulted in a dramatic improvement in life expectancy and in quality of life for people with severe hemophilia. However, the contamination of the blood supply with hepatitis viruses, and subsequently HIV, resulted in widespread transmission of these bloodborne infections within the hemophilia population. The introduction of viral inactivation steps in the preparation of plasma-derived products in the mid-1980s greatly reduced the risk of HIV and hepatitis; the risks were further reduced by the production of recombinant FVIII and FIX proteins in the 1990s. It is uncommon for hemophilic patients born after 1985 to have contracted either hepatitis or HIV, and for these individuals, life expectancy is ~65 years. In fact, since 1998, new infections with viral hepatitis or HIV have not been reported in hemophilia patients.

Factor replacement for hemophilia has been the mainstay of therapy for half a century; however, advances including uniquely functioning molecules and gene therapy have expanded treatment approaches. Factor replacement has been provided either in response to a bleeding episode or as prophylactic treatment. Primary prophylaxis is defined as maintaining the missing clotting factor at levels ~1% or higher on a regular basis in order to prevent bleeds, especially the onset of hemarthroses. Hemophilic boys receiving regular infusions of FVIII (3 days/week) or FIX (2 days/week) can reach puberty without detectable joint abnormalities. Therefore, prophylactic treatment has become more common. The Centers for Disease Control and Prevention reported that more than 51% of children with severe hemophilia who are aged <6 years receive prophylaxis, increasing considerably from 33% in 1995. Although prophylaxis with factor concentrates is the standard care for children and adults with severe hemophilia, teenagers and young adults do not always maintain treatment due to high cost and lifestyle factors including difficulties accessing peripheral veins for two-to-three times a week infusions, and potential infectious and thrombotic risks of long-term central vein catheters.

Treatment of hemophilia bleeds requires the following: (1) prompt initiation of factor replacement as symptoms often precede objective evidence of bleeding, especially for classic symptoms of bleeding into the joint in a reliable patient, headaches, or major trauma; and (2) avoidance of antiplatelet drugs.

FVIII and FIX are dosed in units. One unit is defined as the amount of FVIII (100 ng/mL) or FIX (5 µg/mL) in 1 mL of normal plasma. One unit of FVIII per kilogram of body weight increases the plasma FVIII level by 2%. One can calculate the dose needed to increase FVIII levels to 100% in a 70-kg severe hemophilia patient (<1%) using the simple formula below. Thus, 3500 units of FVIII will raise the circulating level to 100%.

$$\text{FVIII dose (IU)} = \frac{\text{Target FVIII levels} - \text{FVIII baseline levels}}{\times \text{body weight (kg)} \times 0.5 \text{ unit/kg}}$$

The doses for FIX replacement are different from those for FVIII, because FIX recovery after infusion is usually only 50% of the predicted value. Therefore, the formula for FIX replacement is as follows:

$$\text{FIX dose (IU)} = \text{Target FIX levels} - \text{FIX baseline levels} \\ \times \text{body weight (kg)} \times 1 \text{ unit/kg}$$

The FVIII half-life of 8–12 h requires injections twice a day to maintain therapeutic levels, whereas the FIX half-life is longer, ~24 h, so that once-a-day injection is sufficient. In specific situations such as after surgery, continuous infusion of factor may be desirable because of its safety in achieving sustained factor levels at a lower total cost.

Cryoprecipitate is enriched with FVIII protein bound to VWF (each bag contains ~80 IU of FVIII). Because of the risk of blood-borne diseases, this product should be used only in emergencies when factor concentrates are not available, although cryoprecipitate may be the only source of FVIII in developing countries.

Mild bleeds such as uncomplicated hemarthroses or superficial hematomas require achieving an initial factor level of 30–50%. Additional doses to maintain levels of 15–25% for 2 or 3 days are indicated for severe hemarthroses, especially when these episodes affect the “target joint.” Large hematomas, or bleeds into deep muscles, require factor levels of 50% or even higher if the clinical symptoms do not improve, and factor replacement may be required for a period of 1 week or longer. The control of serious bleeds, including those that affect the oropharyngeal spaces, CNS, and the retroperitoneum, requires sustained protein levels of 50–100% for 7–10 days. Prophylactic replacement for surgery is aimed at achieving normal factor levels (100%) for a period of 7–10 days; replacement can then be tapered depending on the extent of the surgical wounds. Oral surgery is associated with extensive tissue damage that usually requires factor replacement for 1–3 days coupled with oral antifibrinolytic drugs.

NONTRANSFUSION THERAPY IN HEMOPHILIA

DDAVP (1-Amino-8-D-Arginine Vasopressin) DDAVP is a synthetic vasopressin analog that causes a transient rise in FVIII and VWF, but not FIX by release from stores in vascular endothelial cells. Patients with moderate or mild hemophilia A should be tested to determine if they respond to DDAVP before use. DDAVP at doses of 0.3 µg/kg body weight, over a 20-min period, is expected to raise FVIII levels by two- to threefold over baseline, peaking between 30 and 60 min after infusion. DDAVP does not improve FVIII levels in severe hemophilia A patients because no stores are available to release. Repeated dosing of DDAVP results in tachyphylaxis as storage pools are depleted. After three consecutive doses, if further therapy is indicated, exogenous FVIII is required.

Antifibrinolytic Drugs Bleeding in the gums, in the gastrointestinal tract, and during oral surgery can be treated with oral antifibrinolytic drugs such as ε-amino caproic acid (EACA) or tranexamic acid to prevent fibrin degradation by plasmin. The duration of the treatment depending on the clinical indication is 1 week or longer. Tranexamic acid is given at doses of 25 mg/kg three to four times a day. EACA treatment requires a loading dose of 200 mg/kg (maximum of 10 g) followed by 100 mg/kg per dose (maximum 30 g/d) every 6 h. These drugs are not indicated to control hematuria because of concern for forming an occlusive clot in the lumen of genitourinary tract structures.

COMPLICATIONS

Inhibitor Formation The formation of alloantibodies to FVIII or FIX is the major complication of hemophilia treatment. The prevalence of inhibitors to FVIII is estimated to be ~30% in severe hemophilia A patients and 10% among patients with nonsevere hemophilia A. Inhibitors to FIX are detected in only 3–5% of all hemophilia B patients. The high-risk group for inhibitor formation includes severe deficiency

(>80% of all cases of inhibitors), familial history of inhibitor, African descent, mutations in the FVIII or FIX gene resulting in deletion of large coding regions, or gross gene rearrangements. Inhibitors usually appear early in life, at a median of 2 years of age, and after 10 cumulative days of exposure. However, intensive replacement therapy such as for major surgery, intracranial bleeding, or trauma increases the risk of inhibitor formation for patients of all ages and degree of clinical severity, such that patients require close laboratory monitoring in the weeks following these events.

The clinical diagnosis of an inhibitor is suspected when patients do not respond to factor replacement at therapeutic doses. Inhibitors increase both morbidity and mortality in hemophilia. Because early detection of an inhibitor is critical to a successful correction of the bleeding or to eradication of the antibody, most hemophilia centers perform annual screening with aPTT and mixing studies. The Bethesda assay uses a similar principle as a mixing study and defines the specificity of the inhibitor and its titer. The results are expressed in Bethesda units (BU), in which 1 BU is the amount of antibody that neutralizes 50% of the FVIII or FIX present in normal plasma after 2 h of incubation at 37°C. Clinically, inhibitor patients are classified as low responders or high responders, with response defined as increase in antibody titer; knowledge of responder type guides therapy. Therapy for inhibitor patients has two goals: the control of acute bleeding episodes and the eradication of the inhibitor. For the control of bleeding episodes, low responders, those with titer <5 BU, respond well to high doses of human FVIII (50–100 U/kg), with minimal or no increase in the inhibitor titers. However, high-responder patients, those with initial inhibitor titer >5 BU or an anamnestic response with increase in the antibody titer to >5 BU, even if low titer initially, do not respond to FVIII. The control of bleeding episodes in high-responder patients can be achieved by using concentrates enriched for prothrombin, FVII, FIX, FX (prothrombin complex concentrates [PCCs] but usually activated PCCs [aPCCs]), and recombinant activated factor VII (FVIIa) known as “bypass agents” as they activate coagulation downstream of the inhibited/absent factor or through a different pathway (Fig. 116-1). For FIX inhibitor patients, high doses of FIX can be used (<5 BU); however, allergic or anaphylactic reactions are common in FIX inhibitor patients; thus bypass products should be used to treat or prevent bleeding as well as for those cases of high titer inhibitors. For eradication of the inhibitory antibody, immunosuppression alone is not effective. The most effective strategy is immune tolerance induction (ITI) based on daily infusion of the missing protein until the inhibitor disappears, typically requiring periods >1 year, with success rates of ~60%. The management of patients with severe hemophilia and inhibitors resistant to ITI is challenging. The use of anti-CD20 monoclonal antibody (rituximab) combined with ITI was thought to be effective but while it reduces the inhibitor titers in some cases, sustained eradication is uncommon.

Other Therapeutic Approaches for Hemophilia A and B Engineered clotting factors, using fusion to polyethylene glycol (FVIII, FIX), IgG1-Fc (FVIII, FIX) or albumin (FIX) with resultant longer half-lives, have been in development, with one currently approved for use. These new-generation products (for FVIII and FIX) aim to facilitate prophylaxis by requiring fewer weekly injections to maintain circulating levels >1%, with infusion frequency decreasing from 3 to 2 days a week in hemophilia A, and notably for hemophilia B, only once-a-week injections of long-acting FIX are required. Other novel approaches to manipulating the coagulation cascade components, such as targeting the natural anticoagulants and inhibitors of activation of coagulation, are in development.

Emicizumab is an asymmetric bispecific antibody with one immunoglobulin variable chain region that binds FIXa and another that binds FX bringing them in close contact resulting in activation of FX by FIXa. FXa subsequently cleaves prothrombin to thrombin—without the need for FVIII (Fig. 116-2). It is effective in patients with severe hemophilia A with or without inhibitors. After initial once-a-week subcutaneous injections (an improvement

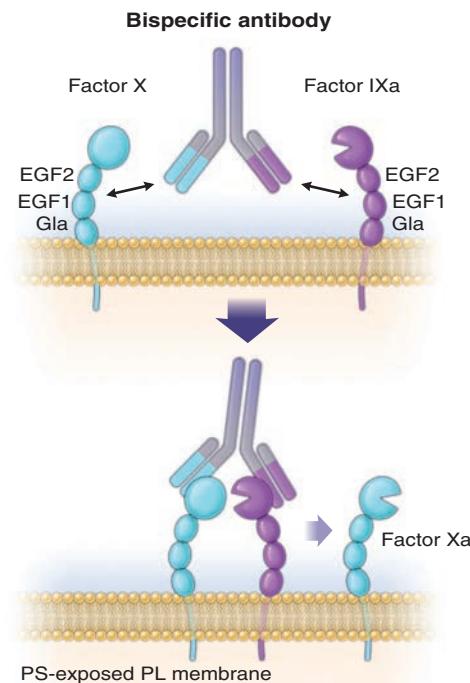


FIGURE 116-2 Mechanism of action of emicizumab. Emicizumab is a bifunctional antibody; the two binding sites recognize different protein sequences, unlike normal antibodies where both variable regions recognize the same antigen. One arm of emicizumab recognizes factor IXa and the other factor X. It functions to bring these two factors in proximity so that factor IXa can activate factor X to factor Xa, which then cleaves prothrombin to thrombin and activates the clotting cascade. (From T Kitazawa, M Shima: Emicizumab, a humanized bispecific antibody to coagulation factors IXa and X with a factor VIIa-cofactor activity. *Int J Hematol* 111:20, 2020.)

over intravenous administration of factors) for 4 weeks, patients can be maintained with once-a-month dosing to prevent spontaneous bleeds, an overwhelmingly dramatic improvement in quality of life when compared to even the twice-weekly infusion schedule of “long-acting” FVIII compounds. Breakthrough bleeds can occur, however, and need to be carefully managed, as a small number of patients with inhibitors treated with aPCC or recombinant FVIIa developed thrombotic events or fatal thrombotic microangiopathy.

These X-linked disorders are ideally suited for gene therapy as small increases in plasma factor level will result in significant clinical improvement. FIX has been the most studied as the gene is smaller and easier to package in the viral vectors used. In one approach, the sequence of a known spontaneous FIX gain of function mutation that has marked increase in specific activity, FIX Padua, is used so that small increments in plasma level of FIX are also accompanied by even greater increase in functional activity. The larger FVIII gene has also been successfully transferred through an adeno-associated viral vector to a few patients with hemophilia A. The early results appear promising. Complications include transaminitis and loss of gene expression for a variety of reasons; no gene therapy approaches have regulatory approval yet (*Chap. 470*).

INFECTIOUS DISEASES

Hemophilia patients treated with clotting factor concentrates before the development of recombinant factors in the 1990s were almost universally infected with hepatitis C virus (HCV) and HIV. These infections are the major cause of morbidity and the second leading cause of death in these patients. Co-infection of HCV and HIV, present in almost 50% of hemophilia patients, is an aggravating factor for the evolution of liver disease as correction of both genetic and acquired (secondary to liver disease) factor deficiencies may be needed. Improvements in treatment of both HIV and HCV have altered the devastating prognosis for many infected patients.

In some select cases with cirrhosis, liver transplant has been performed, which also is curative for hemophilia.

EMERGING CLINICAL PROBLEMS IN AGING HEMOPHILIA PATIENTS

The number of patients living with hemophilia well into adulthood has increased with the advances in treatments. The life expectancy of patients with severe hemophilia is now only ~10 years shorter than the general male population, and near normal in patients with mild or moderate hemophilia. The older hemophilia population has distinct needs relating to more severe arthropathy, chronic pain, and high rates of HCV and/or HIV infections.

Although mortality from coronary artery disease is lower in hemophilia patients with hypocoagulability decreasing thrombus formation, atherogenesis is not prevented. Typical cardiovascular risk factors such as age, obesity, and smoking, along with physical inactivity, hypertension, and chronic renal disease are seen in hemophilia patients as in the general population.

Management of an acute ischemic event and coronary revascularization should include collaboration among hematologists, cardiologists, and internists. Cancer due to HIV- and HCV-related malignancies is also a concern in this population, with hepatocellular carcinoma (HCC) the most common cause of death in HIV-negative patients. The recommendations for cancer screening for the general population should be the same for age-matched hemophilia patients, including routine screening for HCC. Screening for GU or GI tract neoplasms in patients with hematuria or hematochezia may be delayed. Hemophilia patients benefit from the same preventive and therapeutic approaches to minimize the risk of cardiovascular disease and malignancy as the general population.

MANAGEMENT OF CARRIERS OF HEMOPHILIA

Women carriers of hemophilia with factor levels ~50% of normal may not have an increased risk for bleeding. However, a wide range of factor activity (22–116%) due to random inactivation of the X chromosome (*lyonization*) can occur and lead to unexpected bleeding in women with low levels. The factor level of carriers should be measured to optimize perioperative management. During pregnancy, FVIII levels increase approximately two- to threefold compared to nonpregnant women, whereas the FIX increase is less pronounced. After delivery, a rapid fall in the pregnancy-induced rise of maternal clotting factor levels occurs resulting in an imminent risk of bleeding that can be prevented by infusion of factor concentrate to levels of 50–70% for 3 days for vaginal delivery and up to 5 days for cesarean delivery. In mild cases, the use of DDAVP and/or antifibrinolytic drugs is recommended.

FACtOR XI DEFICIENCY

Factor XI deficiency, also known as hemophilia C, is a rare autosomal bleeding disorder that occurs at a frequency of one in a million. However, the disease is highly prevalent among Ashkenazi and Iraqi Jewish populations, reaching a frequency of 6% heterozygotes and 0.1–0.3% homozygotes. More than 65 mutations in the FXI gene have been reported, whereas fewer mutations (two to three) are found among affected Jewish populations.

Normal FXI clotting activity levels range from 70–150 U/dL. Levels vary depending on the presence of heterozygous, homozygous, or double heterozygous mutations with levels <1 U/dL seen in the latter two. Patients with FXI levels <10% of normal have a high risk of bleeding, but the phenotype does not always correlate with FXI clotting activity. The family history is informative, with the bleeding risk based on bleeding in kindreds. Clinically, mucocutaneous hemorrhages such as bruises, gum bleeding, epistaxis, hematuria, and menorrhagia are common, especially following trauma. This hemorrhagic phenotype suggests that tissues rich in fibrinolytic activity are more susceptible to FXI deficiency. Postoperative bleeding is common but not always present, even among patients with very low FXI levels.

FXI replacement is indicated in patients with severe disease for major surgical procedures. A negative history of bleeding complications

following invasive procedures does not exclude the possibility of an increased risk for hemorrhage.

TREATMENT

Factor XI Deficiency

Sources of FXI are limited to FFP in the United States, while a plasma-derived FXI concentrate is available in other countries. FFP at doses of 15–20 mL/kg to maintain trough levels ranging from 10–20% can be given every other day in the setting of bleeding or major surgery as FXI has a half-life of 40–70 h. Antifibrinolytic drugs can be used for minor bleeds and as adjunctive treatment with FXI replacement with the exception of GU tract bleeding. The development of an FXI inhibitor can be seen in 10% of severely FXI-deficient patients. Although inhibitors are not associated with spontaneous bleeding, bleeding with surgery or trauma can be severe; treatment with PCC/apCC or recombinant activated FVII is effective.

RARE BLEEDING DISORDERS

Inherited disorders resulting from deficiencies of clotting factors other than FVIII, FIX, and FXI (Table 116-1) occur infrequently. Bleeding manifestations vary from generally asymptomatic as with dysfibrinogenemia or FVII deficiency to life-threatening as with FX or FXIII deficiency. In contrast to hemophilia, hemarthroses are rare but bleeding in the mucosal tract or after umbilical cord clamping is common. Individuals heterozygous for plasma coagulation deficiencies are often asymptomatic. The laboratory assessment for the specific deficient factor following screening with general coagulation tests (Table 116-1) identifies the diagnosis.

Replacement therapy using FFP or PCCs for deficiencies provides adequate hemostasis for bleeds or prophylactic treatment, although specific concentrates for FX and fibrinogen are available. Cryoprecipitate or FXIII concentrate is needed for FXIII deficiency. FVII deficiency, like FXI, has an increased prevalence in the Ashkenazi Jewish population and is best treated with rVIIa rather than FFP or PCCs depending on the severity of bleeding or type of surgery.

FAMILIAL MULTIPLE COAGULATION DEFICIENCIES

Several bleeding disorders are characterized by the inherited deficiency of more than one plasma coagulation factor. To date, the genetic defects in two of these diseases have been characterized, and they provide new insights into the regulation of hemostasis by gene-encoding proteins outside blood coagulation.

Combined Deficiency of FV and FVIII Patients with combined FV and FVIII deficiency exhibit ~5% of residual clotting activity of each factor, yet it is associated with a mild bleeding tendency, often following trauma. A mutation in the lectin mannose binding 1 (*LMAN1*) gene, a mannose-binding protein localized in the Golgi apparatus that functions as a chaperone for both FV and FVIII, is responsible. In other families, mutations in the multiple coagulation factor deficiency 2 (*MCFD2*) gene have been defined; this gene encodes a protein that forms a Ca^{2+} dependent complex with *LMAN1* and provides cofactor activity in the intracellular mobilization of both FV and FVIII. Replacement therapy to control or prevent bleeding consists of FFP to maintain FV levels and DDAVP or FVIII concentrate to achieve FVIII levels of 20–40%. Alternatively, platelets, which contain FV, can also be used.

Multiple Deficiencies of Vitamin K-Dependent Coagulation Factors

Two enzymes involved in vitamin K metabolism have been associated with combined deficiency of all vitamin K-dependent proteins, including the procoagulant proteins prothrombin (II), VII, IX, and X and the anticoagulant proteins C and S. Vitamin K is a fat-soluble vitamin that is a cofactor for carboxylation of the gamma carbon of the glutamic acid residues in the vitamin K-dependent factors, a critical step for calcium and phospholipid binding of these proteins.

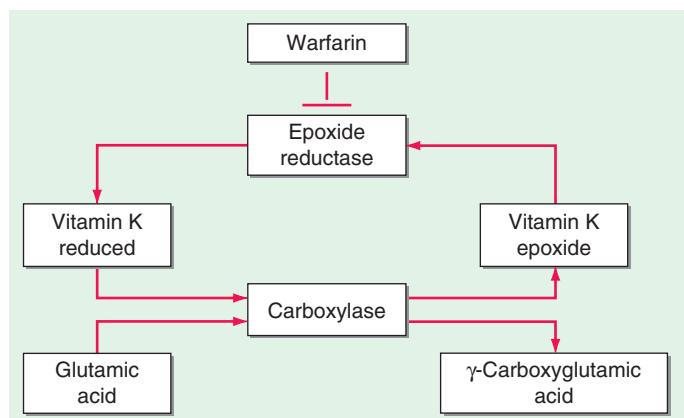


FIGURE 116-3 The vitamin K cycle. Vitamin K is a cofactor for the formation of γ -carboxyglutamic acid residues on coagulation proteins. Vitamin K-dependent γ -glutamylcarboxylase, the enzyme that catalyzes the vitamin K epoxide reductase, regenerates reduced vitamin K. Warfarin blocks the action of the reductase and competitively inhibits the effects of vitamin K.

(Fig. 116-3). The enzymes γ -glutamylcarboxylase and epoxide reductase are critical for the metabolism and regeneration of vitamin K. Mutations in the genes encoding the γ -carboxylase (GGCX) or vitamin K epoxide reductase complex 1 (VKORC1) result in defective enzymes and thus in vitamin K-dependent factors with reduced activity, varying from 1–30% of normal. Patients can have mild to severe bleeding episodes present from birth. Some patients respond to oral vitamin K1 (5–20 mg/d), or parenteral vitamin K1 at doses of 5–20 mg/week. For severe bleeding, replacement therapy with PCC may be necessary.

DISSEMINATED INTRAVASCULAR COAGULATION

In 2001, the International Society on Thrombosis and Haemostasis (ISTH) defined disseminated intravascular congestion (DIC) as “an acquired syndrome characterized by the intravascular activation of coagulation with loss of localization arising from different causes that can originate from and cause damage to the microvasculature, which if sufficiently severe, can produce organ dysfunction.” Many disparate processes are associated with DIC (Table 116-2).

The most common causes are bacterial sepsis, although viral and fungal sepsis can also cause DIC, trauma, obstetric causes such as abruptio placentae or amniotic fluid embolism, and malignant disorders especially mucin-producing adenocarcinomas and acute promyelocytic leukemia. Activation of inflammatory pathways in response to infectious pathogens results in increased expression of tissue factor, activation of neutrophils and monocytes with release of cytokines and development of neutrophil extracellular traps, and release of polyphosphates that engage in cross talk with the coagulation system to cause thrombin generation; this process is known as *thrombo-inflammation*. Damage to vascular endothelial cells results in the loss of their native antithrombotic properties; such damage especially occurs with sepsis and trauma. Systemic inflammatory response syndrome (SIRS) and cytokine storm are cytokine-mediated exuberant inflammatory responses often in the setting of infection that are associated with increased mortality and DIC. Purpura fulminans is a severe form of DIC resulting in thrombosis of extensive areas of the skin; it affects predominantly young children following viral or bacterial infection, particularly those with inherited or acquired hypercoagulability due to deficiencies of the components of the protein C pathway. Neonates homozygous for protein C deficiency can develop neonatal purpura fulminans with or without thrombosis of large vessels.

The central mechanism of DIC is the uncontrolled generation of thrombin by multiple mechanisms (Fig. 116-4). Simultaneous disruption of the physiologic anticoagulant mechanisms and abnormal fibrinolysis further accelerate the process. These abnormalities contribute to systemic fibrin deposition in small and midsize vessels. The duration and intensity of the fibrin deposition can compromise

TABLE 116-2 Common Clinical Causes of Disseminated Intravascular Coagulation

SEPSIS	IMMUNOLOGIC DISORDERS
<ul style="list-style-type: none"> Bacterial: Staphylococci, streptococci, pneumococci, meningococci, gram-negative bacilli Viral Mycotic Parasitic Rickettsial 	<ul style="list-style-type: none"> Acute hemolytic transfusion reaction Organ or tissue transplant rejection Immunotherapy Graft-versus-host disease
TRAUMA AND TISSUE INJURY	DRUGS
<ul style="list-style-type: none"> Brain injury (gunshot) Extensive burns Fat embolism Rhabdomyolysis 	<ul style="list-style-type: none"> Fibrinolytic agents Aprotinin Warfarin (especially in neonates with protein C deficiency) Prothrombin complex concentrates Recreational drugs (amphetamines)
VASCULAR DISORDERS	ENVENOMATION
<ul style="list-style-type: none"> Giant hemangiomas (Kasabach-Merritt syndrome) Large vessel aneurysms (e.g., aorta) 	<ul style="list-style-type: none"> Snake Insects
OBSTETRICAL COMPLICATIONS	LIVER DISEASE
<ul style="list-style-type: none"> Abruptio placenta Amniotic fluid embolism Dead fetus syndrome Septic abortion 	<ul style="list-style-type: none"> Fulminant hepatic failure Cirrhosis Fatty liver of pregnancy
CANCER	MISCELLANEOUS
<ul style="list-style-type: none"> Adenocarcinoma (prostate, pancreas, etc.) Hematologic malignancies (acute promyelocytic leukemia) 	<ul style="list-style-type: none"> Shock Respiratory distress syndrome Massive transfusion

the blood supply of many organs, especially the lung, kidney, liver, and brain, with consequent organ failure; for example, pulmonary microvascular thrombosis is a component of adult respiratory distress syndrome (ARDS). The sustained activation of coagulation and formation of fibrin can result in consumption of clotting factors and platelets,

which in turn leads to systemic bleeding that can be aggravated by secondary hyperfibrinolysis that occurs in late stages of DIC.

Clinical manifestations of DIC are related to the magnitude of the imbalance of hemostasis, to the underlying disease, or to both. The most common clinical findings include petechiae, ecchymoses, and bleeding ranging from oozing from venipuncture sites to severe hemorrhage from the gastrointestinal tract, lung, or into the CNS. In chronic DIC, the bleeding symptoms are discrete and restricted to skin or mucosal surfaces. The hypercoagulability of DIC manifests as the occlusion of vessels in the microcirculation and resulting organ failure. Thrombosis of large vessels and cerebral embolism can also occur. Hemodynamic complications and shock are common among patients with acute DIC, due to the underlying disease, with mortality ranging from 30 to >80%.

Making the diagnosis of DIC can be difficult. The ISTH has developed a validated scoring tool to aid in the diagnosis of overt DIC with a separate tool for pregnant women. It incorporates platelet count, d-dimer level, prothrombin time (PT), and fibrinogen level, and assigns points for different levels of each with the aggregate score helping to make the diagnosis of DIC (Table 116-3). The peripheral smear should be assessed for schistocytes. The laboratory diagnosis of DIC should prompt a search for the underlying disease if not already apparent. In critically ill patients, these tests should be repeated over a period of 6–8 h as patients can rapidly deteriorate.

Chronic DIC Low-grade, compensated DIC can occur in clinical situations including giant hemangioma, metastatic carcinoma, or the dead fetus syndrome. Plasma levels of FDP or d-dimers are elevated. aPTT, PT, and fibrinogen values are within the normal range or high. Mild thrombocytopenia or normal platelet counts are also common findings. Red cell fragmentation is often detected but at a lower degree than in acute DIC.

Differential Diagnosis Distinguishing between DIC and severe liver disease is challenging and requires serial measurements of the laboratory parameters of DIC. Patients with severe liver disease manifest laboratory features including thrombocytopenia due to platelet sequestration, portal hypertension, or hypersplenism; decreased synthesis of coagulation factors and natural anticoagulants; and elevated levels of d-dimer. However, in contrast to DIC, these laboratory parameters in liver disease do not change rapidly.

Although microangiopathic disorders such as acquired thrombotic thrombocytopenic purpura present with acute onset accompanied by

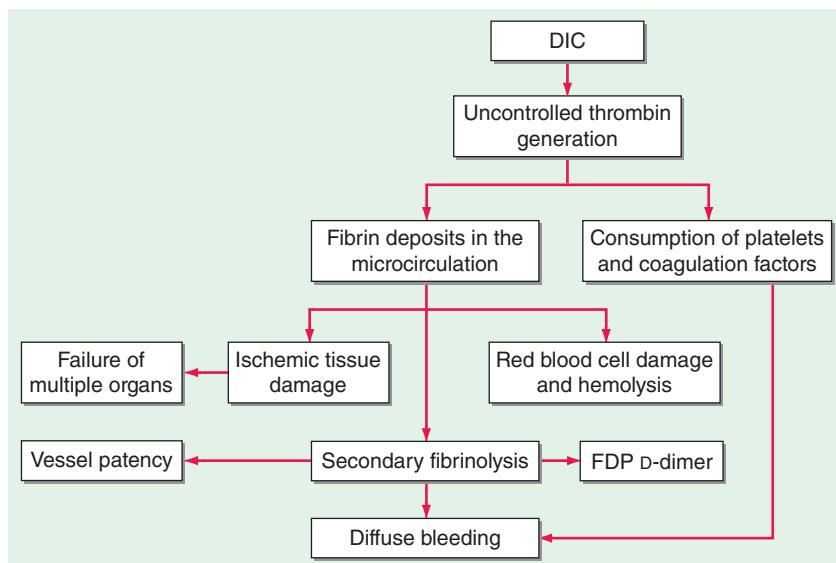


FIGURE 116-4 The pathophysiology of disseminated intravascular coagulation (DIC). Interactions between coagulation and fibrinolytic pathways result in bleeding and thrombosis in the microcirculation in patients with DIC. FDP, fibrin degradation product.

TABLE 116-3 ISTH Criteria for Overt DIC

PARAMETER	VALUE	POINTS
Platelets	>100,000 × 10 ⁹ /L	0
	>50 – <100 × 10 ⁹ /L	1
	<50 × 10 ⁹ /L	2
D-dimer*	Normal	0
	Moderate increase	2
	Severe increase	3
Prothrombin time (PT) prolonged	<3 s	0
	3 – <6 s	1
	>6 s	2
Fibrinogen	>1 g/L	0
	<1 g/L	1
Total Score		<5 Low-grade DIC >5 Overt DIC

*D-dimer assays are not standardized and have different ranges of normal. Check your institution range of normal to assess degree of increase.

Note: A score of <5 suggests non-overt DIC/low-grade DIC and should be repeated every 1–2 days. A score of >5 suggests overt DIC, lab values should be repeated daily to assess critical changes. Not to be used in pregnant patients.

thrombocytopenia, red cell fragmentation, and multiorgan failure, the clinical presentation and laboratory findings such as an inhibitor to ADAMTS13 levels assist in making the microangiopathic disorder diagnosis (**Chap. 115**).

TREATMENT

Disseminated Intravascular Coagulation

The morbidity and mortality associated with DIC are primarily related to the underlying disease. Management of the underlying disease is required to control and eliminate DIC; however, support with platelets and coagulation factors may be needed until the inciting cause is under control. Many patients with overt DIC are critically ill, usually requiring management in the intensive care unit to treat shock physiology and other manifestations of the underlying illness.

MANAGEMENT OF HEMORRHAGIC SYMPTOMS

Patients with active bleeding or at high risk of bleeding during invasive procedures or after chemotherapy require transfusion support; however, transfusion solely to correct mildly to moderately abnormal coagulation parameters is not indicated. Platelet transfusion for platelet counts <10,000–20,000/µL and replacement of fibrinogen and coagulation factors with FFP, with cryoprecipitate or fibrinogen concentrate as a source of fibrinogen, are indicated with amounts determined by the degree of abnormal PT, aPTT, and fibrinogen levels, as well as severity of bleeding or bleeding risk with invasive procedures. For these situations, fibrinogen level should be maintained at >150 mg/dL and PT prolonged no more than 3 s above the upper limit of normal. Vitamin K should be given. Patients should be frequently monitored, and transfusion support adjusted as the patient's condition changes and dictates.

REPLACEMENT OF COAGULATION OR FIBRINOLYSIS INHIBITORS

Anticoagulants such as heparin, antithrombin III (ATIII), and thrombomodulin concentrates, and antifibrinolytic drugs have all been tried in the treatment of DIC. Low doses of continuous-infusion heparin (5–10 U/kg per h) may be effective in patients with low-grade DIC associated with solid tumors, acute promyelocytic leukemia, or in a setting with recognized thrombosis. Heparin is also indicated for the treatment of purpura fulminans, during the surgical resection of giant hemangiomas, and during removal of a dead fetus. In acute hemorrhagic DIC, the use of heparin is likely to aggravate bleeding. The use of heparin in patients with severe DIC, although demonstrating improved coagulation parameters, has not

had a survival benefit; professional society recommendations for use vary widely. Although the use of concentrates of the serine protease inhibitors, antithrombin and thrombomodulin, for sepsis demonstrated little efficacy in all treated patients, post hoc analyses of those with sepsis and confirmed DIC suggest a survival advantage and require further study. Activated protein C treatment for septic shock was withdrawn from the market years ago as findings in clinical practice did not replicate the mortality advantage seen in the clinical trial; impact on DIC was not evaluated.

In patients who have DIC characterized by a primary hyperfibrinolytic state with concomitant severe bleeding, the administration of antifibrinolytics may be considered. However, concern for increasing the risk of thrombosis has led to consideration of concomitant use of heparin. Patients with acute promyelocytic leukemia or those with chronic DIC associated with giant hemangiomas are among the few patients who may benefit from this therapy.

VITAMIN K DEFICIENCY

Vitamin K-dependent proteins are a heterogeneous group, including clotting factor proteins and also proteins found in bone, lung, kidney, and placenta. Vitamin K mediates posttranslational modification of glutamate residues to γ-carboxylglutamate, a critical step for the activity of vitamin K-dependent proteins for calcium binding and proper assembly on phospholipid membranes (Fig. 116-3). Inherited mutations with decreased functional activity of the enzymes GGCX or VKORC1 (see above) result in bleeding disorders. Vitamin K in the diet is often limiting for the carboxylation reaction; thus recycling of the vitamin K by these enzymes is essential to maintain normal levels of vitamin K-dependent proteins. In adults, severe vitamin K deficiency due to low dietary intake in adults is rare but is common in association with the use of broad-spectrum antibiotics, or with disease or surgical interventions that affect the ability of the intestinal tract to absorb vitamin K, through anatomic alterations or by changing the fat content of bile salts and pancreatic enzymes in the proximal small bowel. Chronic liver diseases such as primary biliary cirrhosis also deplete vitamin K stores. Neonatal vitamin K deficiency and the resulting hemorrhagic disease of the newborn have been almost entirely eliminated by routine administration of vitamin K to all neonates. Prolongation of PT values is the most common and earliest finding in vitamin K-deficient patients due to the short half-life of FVII, and occurs before prolongation of the aPTT. Parenteral administration of 10 mg of vitamin K is sufficient to restore normal levels of clotting factor within 8–10 h. More rapid correction of the coagulopathy requires replacement with FFP or PCC, the choice depending on patient intravascular volume status and need for rapidity of correction. The reversal of excessive anticoagulant therapy with vitamin K antagonists, such as warfarin, can be achieved by minimal doses of vitamin K (1 mg orally or by intravenous injection) for asymptomatic patients. This strategy can diminish the risk of bleeding while maintaining therapeutic anticoagulation for an underlying prothrombotic state. For emergent reversal of warfarin in the setting of life-threatening bleeding or need for emergency surgery, use of 4F-PPC is the standard of care.

In patients with underlying vascular disease, vascular trauma, atrial fibrillation, and other comorbidities, re-initiation of anticoagulation needs to be carefully considered to prevent subsequent thromboembolic complications.

COAGULATION DISORDERS ASSOCIATED WITH LIVER FAILURE

The liver is the site of synthesis and clearance of most procoagulant and natural anticoagulant proteins and of essential components of the fibrinolytic system. Liver failure is associated with a high risk of bleeding due to deficient synthesis of procoagulant factors and enhanced fibrinolysis; hepatologists refer to this as accelerated intravascular coagulation and fibrinolysis (AICF). Thrombocytopenia is common in patients with liver disease and may be due to decreased thrombopoietin that is synthesized in the liver, congestive splenomegaly (hypersplenism), or immune-mediated shortened platelet life span

TABLE 116-4 Coagulation Disorders and Hemostasis in Liver Disease

Bleeding	
Portal hypertension	
Esophageal varices	
Thrombocytopenia	
Splenomegaly	
Chronic or acute DIC	
Decreased synthesis of clotting factors	
Hepatocyte failure	
Vitamin K deficiency	
Systemic fibrinolysis	
DIC	
Dysfibrinogenemia	
Thrombosis	
Decreased synthesis of coagulation inhibitors: protein C, protein S, antithrombin	
Hepatocyte failure	
Vitamin K deficiency (protein C, protein S)	
Failure to clear activated coagulation proteins (DIC)	
Dysfibrinogenemia	

Abbreviation: DIC, disseminated intravascular coagulation.

(primary biliary cirrhosis). In addition, several anatomic abnormalities secondary to underlying liver disease further increase the risk of bleeding (Table 116-4). Dysfibrinogenemia is a relatively common finding in patients with liver disease due to impaired fibrin polymerization. The development of DIC in patients with chronic liver disease is not uncommon and may enhance the risk for bleeding. Laboratory evaluation is mandatory for an optimal therapeutic strategy, either to control ongoing bleeding or before invasive procedures. Typically, these patients present with prolonged PT, aPTT, and TT depending on the degree of liver damage, thrombocytopenia, and normal or slight increase in D-dimer. Fibrinogen levels are low only in fulminant hepatitis, decompensated cirrhosis, advanced liver disease, or in the presence of DIC. The presence of prolonged TT and normal fibrinogen and D-dimer levels suggests dysfibrinogenemia. FVIII levels are often normal or elevated in patients with liver failure, and decreased levels suggest superimposed DIC. FV is only synthesized in the hepatocyte and is not a vitamin K-dependent protein; therefore, reduced levels of FV may be an indicator of liver failure. Normal levels of FV and low levels of FVII suggest vitamin K deficiency. Vitamin K levels may be reduced in patients with liver failure due to compromised storage in hepatocellular disease, changes in bile acids, or cholestasis that can diminish the absorption of vitamin K. Replacement with IV vitamin K may improve hemostasis.

Although treatment of bleeding with FFP was the standard approach to correcting hemostasis in patients with liver failure, the use of 4F-PCC is now favored due to lower volume, less increase in portal pressure, reduced risk of circulatory overload, and other complications associated with FFP transfusion. As in any clinical situation, treatment should not be given simply to correct laboratory abnormalities in a patient who is not bleeding or with no need for invasive procedures. Platelet concentrates are indicated when platelet counts are <10,000–20,000/ μ L to control bleeding

or immediately before an invasive procedure if counts are <50,000/ μ L. Cryoprecipitate is indicated only when fibrinogen levels are <100–150 mg/mL unless the patient is bleeding in which case a higher target is used. The use of antifibrinolytic drugs as adjuncts to control bleeding in patients with liver failure is not thought to result in an increased risk of thrombosis; however, their impact on acute thrombosis propagation is not well studied.

Liver Disease and Thromboembolism Bleeding in patients with stable liver disease is often mild or even asymptomatic. However, as the disease progresses, the hemostatic balance is precarious and easily disturbed; comorbid complications such as infections and renal failure can rapidly upset this balance (Fig. 116-5). Past assumptions based on abnormal coagulation tests have been that patients with liver disease have a decreased risk of thrombosis; however, multiple factors contribute to hypercoagulability, including decreased levels of the natural anticoagulant proteins S and C, as well as endothelial cell changes and hemodynamic changes that result in stasis such that portal vein thrombosis is common. Patients with liver disease can also develop deep-vein thrombosis and pulmonary embolism; those with cirrhosis appear to have a 1.5- to 2-fold increase in the rate of venous thromboembolism (VTE). Patients with compensated cirrhosis do not appear to have increased bleeding with the use of VTE prophylaxis or even therapeutic dose heparin to treat acute portal vein thrombosis when carefully managed. In the outpatient setting, warfarin is avoided but low-molecular-weight heparin and direct oral anticoagulants have been safely used to treat thrombosis.

Acquired Inhibitors of Coagulation Factors An acquired inhibitor is an immune-mediated disease characterized by the presence of an autoantibody against a specific clotting factor. Almost half of patients with an acquired factor inhibitor will have an underlying autoimmune or immunoproliferative disorder, malignancy, or be peripartum. FVIII is the most common target of antibody formation and is sometimes referred to as acquired hemophilia A, but inhibitors to prothrombin (FII), FV, FIX, FX, and FXI are also reported. Acquired inhibitor to FVIII occurs predominantly in older adults (median age of 60 years) but occasionally in pregnant or postpartum women with no previous history of bleeding. Bleeding episodes occur commonly in

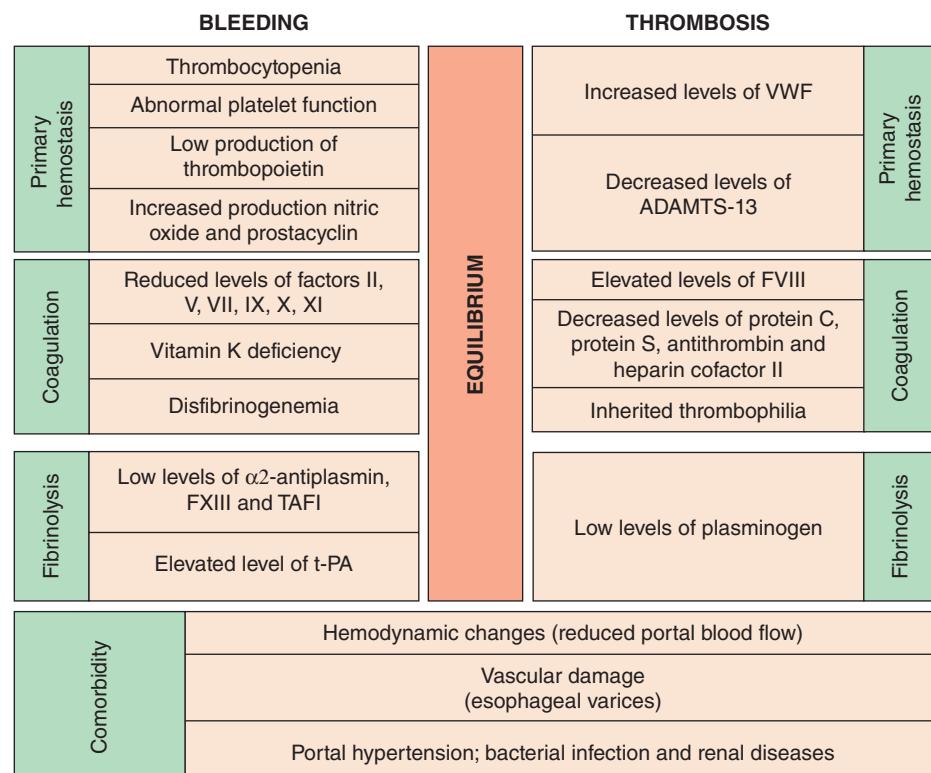


FIGURE 116-5 Balance of hemostasis in liver disease. TAFI, thrombin-activated fibrinolytic inhibitor; t-PA, tissue plasminogen activator; VWF, von Willebrand factor.

soft tissues, the gastrointestinal or urinary tracts, and skin. In contrast to hemophilia, hemarthrosis is rare in these patients. Retroperitoneal hemorrhages and other life-threatening bleeding may appear suddenly. The overall mortality in untreated patients ranges from 8–22%, and most deaths occur within the first few weeks after presentation. The diagnosis is based on the prolonged aPTT with normal PT and TT and a mixing study that does not correct with normal pooled plasma. The Bethesda assay using factor specific-deficient plasma as performed for inhibitor detection in hemophilia will confirm the diagnosis. Treatment of acquired inhibitors of coagulation factors requires control of bleeding and eradication of the inhibitor. Many patients can have life-threatening bleeding. The use of activated “bypass products” such as aPCC or recombinant FVIIa is required. The use of recombinant porcine FVIII can be effective for acquired inhibitors of FVIII. The use of emicizumab to treat acquired FVIII inhibitors has been reported and trials in this population are underway in Europe.

In contrast to hemophilia, inhibitors in nonhemophilic patients are typically responsive to immune suppression, and therapy should be initiated early for most cases. High-dose intravenous γ -globulin and anti-CD20 monoclonal antibody are reported to be effective in patients with autoantibodies to FVIII; however, no firm evidence confirms that these alternatives are superior to the first line of immunosuppressive drugs (glucocorticoids and cyclophosphamide), effective in 70% of patients. Relapse of an inhibitor to FVIII is relatively common (up to 20%) within the first 6 months following withdrawal of immunosuppression; patients should be followed up regularly for relapse.

Topical plasma-derived bovine and human thrombin are commonly used during major cardiovascular, thoracic, neurologic, and pelvic surgeries as well as in trauma patients with extensive burns. Antibody formation to the xenoantigen or its contaminant (bovine clotting protein) has the potential to cross-react with human clotting factors, particularly FV and thrombin and can result in bleeding that can be life-threatening. The development of antibodies to FV with the use of topical preparations of recombinant human thrombin has also been reported. The clinical diagnosis of these acquired coagulopathies is rare but is often complicated by the fact that the bleeding episodes may be detectable during or immediately following major surgery and could be assumed to be due to the procedure itself.

The risk of developing a cross-reacting antibody is increased by repeated exposure to topical thrombin preparations. Thus, a careful medical history of previous surgical interventions that may have occurred even decades earlier is critical to assessing risk.

The laboratory abnormalities include a combined prolongation of the aPTT and PT that often fails to improve by transfusion of FFP and vitamin K, and a mixing study that does not correct with normal pooled plasma. The specificity of the antibody is determined by the measurement of the residual activity of human FV or other suspected human clotting factor. No assays specific for bovine thrombin coagulopathy are commercially available.

No treatment guidelines have been established. Platelet transfusions have been used as a source of FV replacement for patients with FV inhibitors. FFP and vitamin K supplementation may function as co-adjuvants rather than as effective treatments for the coagulopathy itself. Experience with recombinant FVIIa as a bypass agent is limited, and outcomes have been generally poor. Specific treatments to eradicate the antibodies based on immunosuppression with glucocorticoids, intravenous immunoglobulin, or serial plasmapheresis have been sporadically reported. Patients should be advised to avoid any topical thrombin sealant in the future.

The presence of lupus anticoagulant can be associated with venous or arterial thrombotic disease. However, bleeding has also been reported rarely with lupus anticoagulants due to antibodies to prothrombin, resulting in hypoprothrombinemia. Both disorders show a prolonged aPTT that does not correct on mixing. To distinguish acquired inhibitors from lupus anticoagulant, note that the dilute Russell viper venom time (dRVVT) and the hexagonal-phase phospholipids test will be negative in patients with an acquired inhibitor and positive in patients with lupus anticoagulants. Moreover, lupus anticoagulant interferes with the clotting activity of many factors (FVIII, FIX, FXI, FXII), which can be

assessed in the clinical laboratory; acquired inhibitors are specific to a single factor.

ACKNOWLEDGMENT

Valder Arruda and Katherine High wrote this chapter in prior editions and some material from their chapter is included here.

FURTHER READING

- BATTY P, LILLICRAPP D: Advances and challenges for hemophilia gene therapy. *Hum Mol Genet* 28:R95, 2019.
- KITAZAWA T, SHIMA M: Emicizumab, a humanized bispecific antibody to coagulation factors IXa and X with a factor VIIa-cofactor activity. *Int J Hematol* 111:20, 2020.
- LEVI M, SCULLY M: How I treat disseminated intravascular coagulation. *Blood* 131:845, 2018.
- MENEGATTI M et al: Management of rare acquired bleeding disorders. *Hematology Am Soc Hematol Educ Program* 2019:80, 2019.
- O'LEARY JG et al: AGA clinical practice update: coagulation in cirrhosis. *Gastroenterology* 157:34, 2019.
- SRIVASTAVA A et al: WFH guidelines for the management of hemophilia, 3rd edition. *Haemophilia* 26(Suppl 6):1, 2020.

117 Arterial and Venous Thrombosis

Jane E. Freedman, Joseph Loscalzo



OVERVIEW OF THROMBOSIS

GENERAL OVERVIEW

Thrombosis, the obstruction of blood flow due to the formation of clot, may result in tissue anoxia and damage, and it is a major cause of morbidity and mortality in a wide range of arterial and venous diseases and patient populations. As reported in 2020, 655,000 Americans die from heart disease each year, accounting for about 1 in 4 deaths. In 2017, coronary disease killed 365,914 people in the United States, and approximately 805,000 people experienced a heart attack and 795,000 had a stroke.

It is estimated that as many as 600,000 people each year have a pulmonary embolism or deep-venous thrombotic event, and 60,000–80,000 Americans die of these conditions annually. In the nondiseased state, physiologic hemostasis reflects a delicate interplay between factors that promote and inhibit blood clotting, favoring the former. This response is crucial as it prevents uncontrolled hemorrhage and exsanguination following injury. In specific settings, the same processes that regulate normal hemostasis can cause pathologic thrombosis, leading to arterial or venous occlusion. Importantly, many commonly used therapeutic interventions may also alter the thrombotic–hemostatic balance adversely.

Hemostasis and thrombosis primarily involve the interplay among three factors: the vessel wall, coagulation and fibrinolytic proteins, and platelets. Many prevalent acute vascular diseases are due to thrombus formation within a vessel, including myocardial infarction, thrombotic cerebrovascular events, and venous thrombosis. Although the end result is vessel occlusion and tissue ischemia, the pathophysiologic processes governing these pathologies have similarities as well as distinct differences. While many of the pathways regulating thrombus formation are similar to those that regulate hemostasis, the processes triggering or perpetuating thrombosis may be distinct and can vary in different clinical and genetic settings. In venous thrombosis, primary hypercoagulable states reflecting defects in the proteins governing coagulation and/or fibrinolysis or secondary hypercoagulable states involving abnormalities of blood vessels and blood flow or stasis lead to

thrombosis. By contrast, arterial thrombosis is highly dependent on the state of the vessel wall, the platelet, and factors related to blood flow.

ARTERIAL THROMBOSIS

OVERVIEW OF ARTERIAL THROMBOSIS

In arterial thrombosis, platelets and abnormalities of the vessel wall typically play a key role in vessel occlusion. Arterial thrombus forms via a series of sequential steps in which platelets adhere to the vessel wall, additional platelets are recruited, and thrombin is activated (Fig. 117-1). The regulation of platelet adhesion, activation, aggregation, and recruitment will be described in detail below. In addition, while the primary function of platelets is regulation of hemostasis, our understanding of their role in other processes, such as immunity, metastasis, wound healing, and inflammation, continues to evolve.

ARTERIAL THROMBOSIS AND VASCULAR DISEASE

Arterial thrombosis is a major cause of morbidity and mortality both in the United States and, increasingly, worldwide. Although the rates have declined in the United States, the overall burden remains high. Overall, in 2020, heart disease was estimated to cause about 1 of every 4 deaths in the United States. In addition to the 605,000 Americans who will have a new coronary event annually, an additional 200,000 myocardial infarctions occur in those with previous heart attacks. Although the rate of strokes has fallen, each year about 795,000 people experience a new or recurrent ischemic stroke. In 2018, about 1 in 6 deaths from cardiovascular disease were due to stroke in the United States.

THE PLATELET

Many processes in platelets have parallels with other cell types, such as the presence of specific receptors and signaling pathways; however, unlike most cells, platelets lack a nucleus and are unable to adapt to changing biologic settings by altered gene transcription. Platelets sustain limited protein synthetic capacity from megakaryocyte-derived and intracellularly transported messenger RNA (mRNA) and microRNA (miRNA). Most of the molecules needed to respond to various stimuli, however, are maintained in storage granules and membrane compartments.

Platelets are disc-shaped, very small, anucleate cells (1–5 μm in diameter) that circulate in the blood at concentrations of 200–400,000/ μL , with an average life span of 7–10 days. Platelets are derived from megakaryocytes, polyploid hematopoietic cells found in the bone marrow. The primary regulator of platelet formation is thrombopoietin (TPO). The precise mechanism by which megakaryocytes produce and release fully formed platelets is unclear, but the process likely involves formation of proplatelets, pseudopod-like structures generated by the evagination of the cytoplasm from which platelets bud. After release into the circulation, (young, large) platelets may continue to divide. Platelet granules are synthesized in megakaryocytes before thrombopoiesis and contain an array of prothrombotic, proinflammatory, and antimicrobial mediators. The two major types of platelet granules, alpha and dense, are distinguished by their size, abundance, and content. Alpha-granules contain soluble coagulation proteins, adhesion molecules, growth factors, integrins, cytokines, and inflammatory modulators. Platelet dense-granules are smaller than alpha-granules

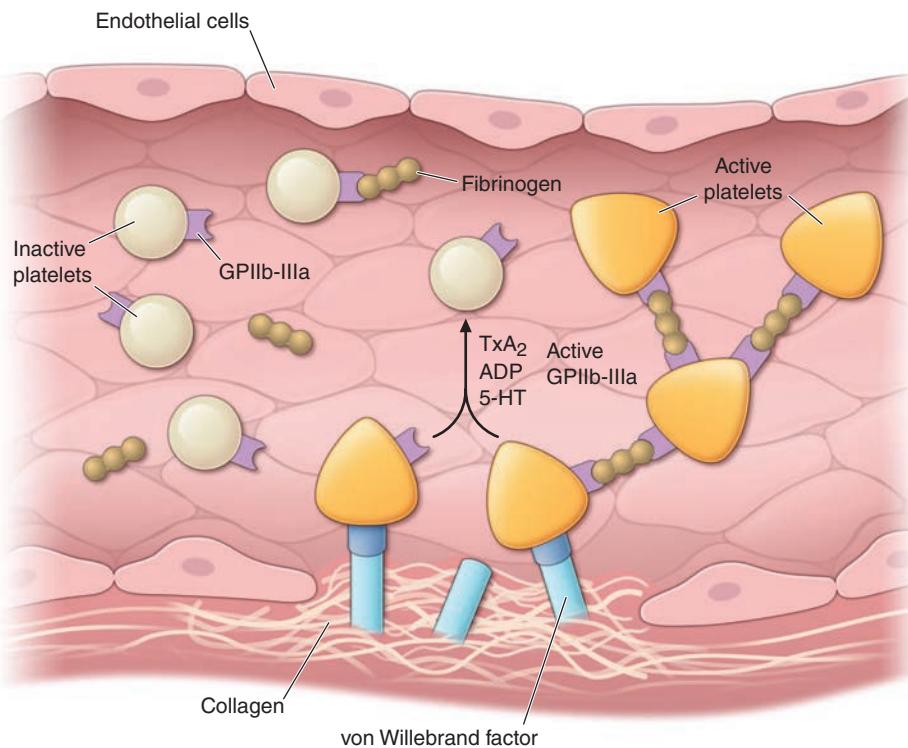


FIGURE 117-1 Platelet activation and thrombosis. Platelets circulate in an inactive form in the vasculature. Damage to the endothelium and/or external stimuli activates platelets that adhere to the exposed subendothelial von Willebrand factor and collagen. This adhesion leads to activation of the platelet, shape change, and the synthesis and release of thromboxane (TxA_2), serotonin (5-HT), and adenosine diphosphate (ADP). Platelet stimuli cause conformational change in the platelet integrin glycoprotein (GP) IIb/IIIa receptor, leading to the high-affinity binding of fibrinogen and the formation of a stable platelet thrombus.

and less abundant. Whereas alpha-granules contain proteins that may be more important in the inflammatory response, dense-granules contain high concentrations of small molecules, including adenosine diphosphate (ADP) and serotonin, that influence platelet aggregation and other related vascular processes, such as vasomotor tone.

Platelet Adhesion (See Fig. 117-1) The formation of a thrombus is initiated by the adherence of platelets to the damaged vessel wall. Damage exposes subendothelial components responsible for triggering platelet reactivity, including collagen, von Willebrand factor, fibronectin, and other adhesive proteins, such as vitronectin and thrombospondin. The hemostatic response may vary, depending on the extent of damage, the specific proteins exposed, and flow conditions. Certain proteins are expressed on the platelet surface that subsequently regulate collagen-induced platelet adhesion, particularly under flow conditions, and include glycoprotein (GP) IV, GPVI, and the integrin $\alpha_2\beta_1$. The platelet GPIb-IX-V complex adhesive receptor is central both to platelet adhesion and to the initiation of platelet activation. Damage to the blood vessel wall exposes subendothelial von Willebrand factor and collagen to the circulating blood. The GPIb-IX-V complex binds to the exposed von Willebrand factor, causing platelets to adhere (Fig. 117-1). In addition, the engagement of the GPIb-IX-V complex with ligand induces signaling pathways that lead to platelet activation. von Willebrand factor-bound GPIb-IX-V promotes a calcium-dependent conformational change in the GPIIb/IIIa receptor, transforming it from an inactive low-affinity state to an active high-affinity receptor for fibrinogen.

Platelet Activation The activation of platelets is controlled by a variety of surface receptors that regulate various functions in the activation process. Platelet receptors control many distinct processes and are stimulated by a wide variety of agonists and adhesive proteins that result in variable degrees of activation. In general terms, the stimulation of platelet receptors triggers two specific processes: (1) activation of internal signaling pathways that lead to further platelet activation and granule release, and (2) the capacity of the platelet to bind to other

adhesive proteins/platelets. Both of these processes contribute to the formation of a thrombus. Stimulation of nonthrombotic receptors results in platelet adhesion or interaction with other vascular cells, including endothelial cells, neutrophils, and mononuclear cells.

Many families and subfamilies of receptors are found on platelets that regulate a variety of platelet functions. These include the seven transmembrane receptor family, which is the main agonist-stimulated receptor family. Several seven transmembrane receptors are found on platelets, including the ADP receptors, prostaglandin receptors, lipid receptors, and chemokine receptors. Receptors for thrombin comprise the major seven transmembrane receptors found on platelets. Among this last group, the first identified was the protease activation receptor 1 (PAR1). The PAR class of receptors has a distinct mechanism of activation that involves specific cleavage of the N-terminus by thrombin, which, in turn, acts as a ligand for the receptor. Other PAR receptors are present on platelets, including PAR2 (not activated by thrombin) and PAR4. Adenosine receptors are responsible for transduction of ADP-induced signaling events, which are initiated by the binding of ADP to purinergic receptors on the platelet surface. There are several distinct ADP receptors, classified as P2X₁, P2Y₁, and P2Y₁₂. The activation of both the P2Y₁₂ and P2Y₁ receptors is essential for ADP-induced platelet aggregation. The thienopyridine derivatives, clopidogrel and prasugrel, are clinically used inhibitors of ADP-induced platelet aggregation.

Platelet Aggregation Activation of platelets results in a rapid series of signal transduction events, including tyrosine kinase, serine/threonine kinase, and lipid kinase activation. In unstimulated platelets, the major platelet integrin GPIIb/IIIa is maintained in an inactive conformation and functions as a low-affinity adhesion receptor for fibrinogen. This integrin is unique as it is only expressed on platelets. After stimulation, the interaction between fibrinogen and GPIIb/IIIa forms intercellular connections between platelets, leading to the formation of a platelet aggregate (Fig. 117-1). A calcium-sensitive conformational change in the extracellular domain of GPIIb/IIIa enables the high-affinity binding of soluble plasma fibrinogen as a result of a complex network of inside-out signaling events. The GPIIb/IIIa receptor serves as a bidirectional conduit with GPIIb/IIIa-mediated signaling (outside-in) occurring immediately after the binding of fibrinogen. This leads to additional intracellular signaling that further stabilizes the platelet aggregate and transforms platelet aggregation from a reversible to an irreversible process (inside-out).

THE ROLE OF PLATELETS AND THROMBOSIS IN INFLAMMATION

Inflammation plays an important role during the acute thrombotic phase of acute coronary and other vascular occlusive syndromes. In the setting of acute upper respiratory infections, people are at higher risk of myocardial infarction and thrombotic stroke. Patients with acute coronary syndromes have not only increased interactions between platelets (homotypic aggregates), but also increased interactions between platelets and leukocytes (heterotypic aggregates) detectable in circulating blood. These latter aggregates form when platelets are activated, often directly by pathogens, and adhere to circulating leukocytes as part of their contribution to the immune process. Platelets bind via P-selectin (CD62P) expressed on the surface of activated platelets to the leukocyte receptor, P-selectin glycoprotein ligand 1 (PSGL-1). This association leads to increased expression of CD11b/CD18 (Mac-1) on leukocytes, which amplifies immunity but may also support further interactions with platelets partially via bivalent fibrinogen linking this integrin with its platelet surface counterpart, GPIIb/IIIa. Platelet surface P-selectin also induces the expression of tissue factor on monocytes, which promotes fibrin formation.

In addition to platelet-monocyte aggregates, the immunomodulator, soluble CD40 ligand (CD40L or CD154), also reflects a link between thrombosis and inflammation. The CD40 ligand is a trimeric transmembrane protein of the tumor necrosis factor family and, with its receptor CD40, is an important contributor to the inflammatory process leading both to thrombosis and atherosclerosis. While many

immunologic and vascular cells have been found to express CD40 and/or CD40 ligand, in platelets, CD40 ligand is rapidly translocated to the surface after stimulation and is upregulated in the newly formed thrombus. The surface-expressed CD40 ligand is cleaved from the platelet to generate a soluble fragment (soluble CD40 ligand).

Links have also been established among platelets, infection, immunity, and inflammation. Bacterial and viral infections are associated with a transient increase in the risk of acute thrombotic events, such as acute myocardial infarction and stroke. In addition, platelets contribute significantly to the pathophysiology and high mortality rates of sepsis. The expression, functionality, and signaling pathways of Toll-like receptors (TLRs) have been established in platelets. Stimulation of platelet TLR2, TLR3, and TLR4 directly and indirectly activates the platelet's thrombotic and inflammatory responses, and live bacteria induce a proinflammatory response in platelets in a TLR2-dependent manner, suggesting a mechanism by which specific bacteria and bacterial components can directly activate platelet-dependent thrombosis. Additionally, viruses, such as SARS-CoV-2, HIV, hepatitis C virus, and Dengue, are also known to cause elevated levels of thrombosis, and recently, platelets have been shown to regulate immune responses to viruses via receptors TLR7 and TLR8.

Risk Factors for Arterial Thrombosis In addition to immune burden, various factors increase the risk of developing arterial thrombosis. Classically, the cardiovascular-dependent risk factors implicated in thrombosis have been hypertension, high levels of low-density lipoprotein cholesterol, and smoking. However, diabetes, pregnancy, age, and chemotherapeutic agents may also contribute to arterial thrombosis. Stillbirth and loss of multiple pregnancies may increase the risk of ischemic stroke and myocardial infarction as does hormonal replacement therapy. Systemic lupus erythematosus and rheumatoid arthritis are now well-recognized risks for thrombosis, and the former, in particular, may contribute in the pediatric population. The anti-phospholipid syndrome is also another widely recognized autoimmune prothrombotic risk for arterial (and venous) thrombosis.

GENETICS OF ARTERIAL THROMBOSIS

Some studies have associated arterial thrombosis with genetic variants (**Table 117-1A**); however, the associations have been weak and not confirmed in larger series. Platelet count and mean platelet volume have been studied by genome-wide association studies (GWAS), and this approach identified signals located to noncoding regions. Of 15 quantitative trait loci associated with mean platelet volume and platelet count, one located at 12q24 is also a risk locus for coronary artery disease.

In the area of genetic variability and platelet function, studies have primarily dealt with pharmacogenetics, the field of pharmacology dealing with the interindividual variability in drug response based on genetic determinants (**Table 117-2**). This focus has been driven by the wide variability among individuals in terms of response to antithrombotic drugs and the lack of a common explanation for this variance. The best described is the issue of "aspirin resistance," although heterogeneity for other antithrombotics (e.g., clopidogrel) has also been extensively examined. Primarily, platelet-dependent genetic determinants have been defined at the level of (1) drug effect, (2) drug compliance, and (3) drug metabolism. Many candidate platelet genes have been studied for their interaction with antiplatelet and antithrombotic agents.

Many patients have an inadequate response to the inhibitory effects of aspirin. Heritable factors contribute to the variability; however, ex vivo tests of residual platelet responsiveness after aspirin administration have not provided firm evidence for a pharmacogenetic interaction between aspirin and *COX1* or other relevant platelet receptors. As such, currently, there is no clinical indication for genotyping to optimize aspirin's antiplatelet efficiency. For the platelet P2Y12 receptor inhibitor clopidogrel, additional data suggest that genetics may affect the drug's responsiveness and utility. The responsible genetic variant appears not to be the expected P2Y12 receptor but an enzyme responsible for drug metabolism. Clopidogrel is a prodrug, and liver metabolism by specific

TABLE 117-1 Heritable Causes of Arterial and Venous Thrombosis**A. Arterial Thrombosis***Platelet Receptors*

- β3 and α2 integrins
- P₁A2 polymorphism
- Fc(gamma)RIIA
- GPIV T13254C polymorphism
- GPIb
- Thrombin receptor PAR1-5061 → D

Redox Enzymes

- Plasma glutathione peroxidase, GPx3, promoter haplotype H2
- H2 promoter haplotype
- Endothelial nitric oxide synthase
-786T/C, -922A/G, -1468T/A
- Paraoxonase
-107T allele, 192R allele

Homocysteine

- Cystathionine β-synthase 833T → C
- 5,10-Methylene tetrahydrofolate reductase (MTHFR) 677C → T

B. Venous Thrombosis*Procoagulant Proteins*

- Fibrinogen
-455G/A, -854G/A
- Prothrombin (20210G → A)

Protein C Anticoagulant Pathway

- Factor V Leiden: 1691G → A (Arg506Gln)
- Thrombomodulin 1481C → T (Ala455Val)

Fibrinolytic Proteins with Known Polymorphisms

- Tissue plasminogen activator (tPA)
7351C/T, 20 099T/C in exon 6, 27 445T/A in intron 10
- Plasminogen activator inhibitor (PAI-1)
4G/5G insertion/deletion polymorphism at position -675

Homocysteine

- Cystathionine β-synthase 833T → C
- 5,10-MTHFR 677C → T

VENOUS THROMBOSIS**OVERVIEW OF VENOUS THROMBOSIS**

Coagulation is the process by which thrombin is activated and soluble plasma fibrinogen is converted into insoluble fibrin. These steps account for both normal hemostasis and the pathophysiologic processes influencing the development of venous thrombosis. The primary forms of venous thrombosis are deep-vein thrombosis (DVT) in the extremities and the subsequent embolization to the lungs (pulmonary embolism [PE]), referred to together as venous thromboembolic disease (VTE). Although the majority of venous thromboembolic events occur as PE or DVT of the lower extremities, up to 10% of events may occur in other vascular locations. Venous thrombosis occurs due to heritable causes (Table 117-1B) and acquired causes (Table 117-3).

DEEP-VENOUS THROMBOSIS AND PULMONARY EMBOLISM

It is estimated that DVT or PE occurs in ~1–2 individuals per 1000 each year, resulting in 300,000–600,000 new cases of VTE each year in the United States. Approximately, 60,000–80,000 deaths are attributed to DVT or PE annually. Of new cases, up to 30% of patients die within 30 days and one-fifth suffer sudden death due to PE; 30% go on to develop recurrent VTE within 10 years. Data from the Atherosclerosis Risk in Communities (ARIC) study reported a 9% 28-day fatality rate from DVT and a 15% fatality rate from PE. PE in the setting of cancer has a 25% fatality rate. The mean incidence of first DVT in the general population is 5 per 10,000 person-years; the incidence is similar in males and females when adjusting for factors related to reproduction and birth control and increases dramatically with age from 2–3 per 10,000 person-years at 30–49 years of age to 20 per 10,000 person-years at 70–79 years of age.

OVERVIEW OF THE COAGULATION CASCADE AND ITS ROLE IN VENOUS THROMBOSIS

Coagulation is defined as the formation of fibrin by a series of linked enzymatic reactions in which each reaction product converts the subsequent inactive zymogen into an active serine protease (Fig. 117-2). This coordinated sequence is called the coagulation cascade and is a key mechanism for regulating hemostasis. Central to the function of the coagulation cascade is the principle of amplification: due to a series of linked enzymatic reactions, a small stimulus can lead to much greater quantities of fibrin, the end product that prevents hemorrhage at the site of vascular injury. In addition to the known risk factors relevant to hypercoagulopathy, stasis, and vascular dysfunction, newer areas of research have identified contributions from procoagulant microparticles, inflammatory cells, microvesicles, and fibrin structure.

The coagulation cascade is primarily initiated by vascular injury exposing tissue factor to blood components (Fig. 117-2). Tissue factor may also be found in bloodborne cell-derived microparticles and, under pathophysiologic conditions, in leukocytes or platelets. Plasma

cytochrome P450 enzymes is required for activation. The genes encoding the CYP-dependent oxidative steps are polymorphic, and carriers of specific alleles of the CYP2C19 and CYP3A4 loci have increased platelet aggregability. Increased platelet activity has also been specifically associated with the CYP2C19*2 allele, which causes loss of platelet function in select patients. Because these are common genetic variants, this observation has been shown to be clinically relevant in large studies. In summary, although the loss-of-function polymorphism in CYP2C19 is the strongest individual variable affecting pharmacokinetics and antiplatelet response to clopidogrel, it only accounts for 5–12% of the variability in ADP-induced platelet aggregation on clopidogrel. In addition, genetic variables do not appear to contribute significantly to the clinical outcomes of patients treated with the P2Y12 receptor antagonists prasugrel or ticagrelor.

TABLE 117-2 Genetic Variation and Pharmacogenetic Responses to Platelet Inhibitors

POTENTIAL GENE ALTERED	TARGET THERAPEUTIC CLASS	SPECIFIC DRUG
<i>P2Y1</i> and <i>P2Y12</i> , <i>CYP2C19</i> , <i>CYP3A4</i> , <i>CYP3A5</i>	ADP receptor inhibitors	Clopidogrel, prasugrel
<i>COX1</i> , <i>COX2</i>	Cyclooxygenase inhibitors	Aspirin
<i>PIA1/A2</i>	Receptor inhibitors	Abciximab, eptifibatide, tirofiban
<i>INTB3</i> , <i>GPIbA</i>	Glycoprotein IIb-IIIa receptor inhibitors	

TABLE 117-3 Acquired Causes of Venous Thrombosis

Surgery
Neurosurgery
Major abdominal surgery
Malignancy
Antiphospholipid syndrome
Other
Trauma
Pregnancy
Long-distance travel
Obesity
Oral contraceptives/hormone replacement
Myeloproliferative disorders
Polycythemia vera

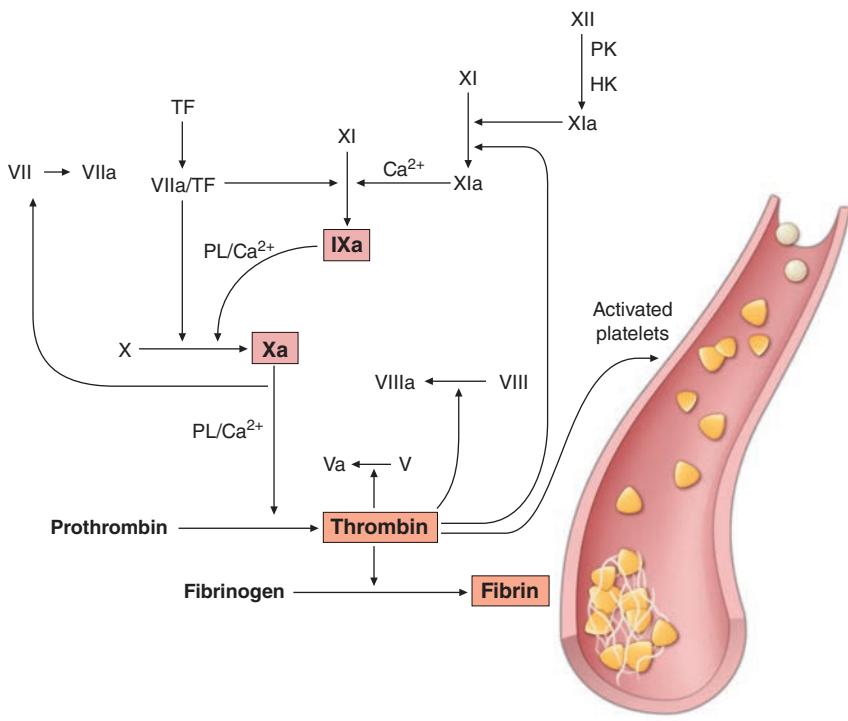


FIGURE 117-2 Summary of the coagulation pathways. Specific coagulation factors ("a" indicates activated form) are responsible for the conversion of soluble plasma fibrinogen into insoluble fibrin. This process occurs via a series of linked reactions in which the enzymatically active product subsequently converts the downstream inactive protein into an active serine protease. In addition, the activation of thrombin leads to stimulation of platelets. HK, high-molecular-weight kininogen; PK, prekallikrein; TF, tissue factor.

factor VII (FVII) is the ligand for and is activated (FVIIa) by binding to tissue factor exposed at the site of vessel damage. The binding of FVII/FVIIa to tissue factor activates the downstream conversion of factor X (FX) to active FX (FXa). In an alternative reaction, the FVII/FVIIa–tissue factor complex initially converts FIX to FIXa, which then activates FX in conjunction with its cofactor factor VIII (FVIIIa). Factor Xa with its cofactor FVa converts prothrombin to thrombin, which then converts soluble plasma fibrinogen to insoluble fibrin, leading to clot or thrombus formation. Thrombin also activates FXIII to FXIIIa, a transglutaminase that covalently cross-links and stabilizes the fibrin clot. Formation of thrombi is affected by mechanisms governing fibrin structure and stability, including specific fibrinogen variants and how they alter fibrin formation, strength, and structure.

Several antithrombotic factors also regulate coagulation; these include antithrombin, tissue factor pathway inhibitor (TFPI), heparin cofactor II, and protein C/protein S. Under normal conditions, these factors limit the production of thrombin to prevent the perpetuation of coagulation and thrombus formation. Typically, after the clot has caused occlusion at the damaged site and begins to expand toward adjacent uninjured vessel segments, the anticoagulant reactions governed by the normal endothelium become pivotal in limiting the extent of this hemostatically protective clot.

RISK FACTORS FOR VENOUS THROMBOSIS

An array of different factors contributes to the risk of VTE, and it is notable that women and men of all ages, races, and ethnicities are at risk for VTE. The risk factors for venous thrombosis are primarily related to hypercoagulability, which can be genetic (Table 117-1) or acquired, or due to immobilization and venous stasis. Independent predictors for recurrence include increasing age, obesity, malignant neoplasm, and acute extremity paresis. It is estimated that 5–8% of the U.S. population has a genetic risk factor known to predispose to venous thrombosis. Often, multiple risk factors are present in a single individual. Significant risk is incurred by major orthopedic, abdominal, or neurologic surgeries. Cancer patients have an approximately fourfold increased risk of VTE as compared with the general population, and cancer patients with VTE have reduced survival.

Hospitalized patients have a greatly increased risk of venous thrombosis with risk factors (increased age, male, ethnicity) and comorbid conditions, including infection, renal disease, and weight loss. Community- or hospital-acquired infection is also associated with increased risk of VTE. Supportive of this, nearly 20% of hospitalized COVID-19 patients are noted to have coagulation abnormalities as well as increased PE, DVT, and peripheral thrombotic risk. Moderate risk is promoted by prolonged bedrest, certain types of cancer, pregnancy, hormone replacement therapy or oral contraceptive use, and other sedentary conditions such as long-distance plane travel. It has been reported that the risk of developing a venous thromboembolic event doubles after air travel lasting 4 h, although the absolute risk remains low (1 in 6000). The relative risk of VTE among pregnant or postpartum women is 4.3, and the overall incidence (absolute risk) is 199.7 per 100,000 woman-years.

■ GENETICS OF VENOUS THROMBOSIS

(See Table 117-2) Less common causes of venous thrombosis are those due to genetic variants. These abnormalities include loss-of-function mutations of endogenous anticoagulants as well as gain-of-function mutations of procoagulant proteins. Heterozygous antithrombin deficiency and homozygosity of the factor V Leiden mutation significantly increase the risk of venous thrombosis. While homozygous protein C or protein S deficiencies are rare and may

lead to fatal purpura fulminans, heterozygous deficiencies are associated with a moderate risk of thrombosis. Activated protein C impairs coagulation by proteolytic degradation of FVa. Patients resistant to the activity of activated protein C may have a point mutation in the FV gene located on chromosome 1, a mutant denoted factor V Leiden. Mildly increased risk has been attributed to elevated levels of procoagulant factors, as well as low levels of tissue factor pathway inhibitor. Polymorphisms of methylene tetrahydrofolate reductase as well as hyperhomocysteinemia have been shown to be independent risk factors for venous thrombosis, as well as arterial vascular disease; however, many of the initial descriptions of genetic variants and their associations with thromboembolism are being questioned in larger, more contemporary studies.

FIBRINOLYSIS AND THROMBOSIS

Specific abnormalities in the fibrinolytic system have been associated with enhanced thrombosis. Factors such as elevated levels of tissue plasminogen activator (tPA) and plasminogen activator inhibitor type 1 (PAI-1) have been associated with decreased fibrinolytic activity and an increased risk of arterial thrombotic disease. Specific genetic variants have been associated with decreased fibrinolytic activity, including the 4G/5G insertion/deletion polymorphism in the *PAI-1* gene. Additionally, the 311-bp Alu insertion/deletion in tPA's intron 8 has been associated with enhanced thrombosis; however, genetic abnormalities have not been associated consistently with altered function or tPA levels, raising questions about the relevant pathophysiologic mechanism. Thrombin-activatable fibrinolysis inhibitor (TAFI) is a carboxypeptidase that regulates fibrinolysis; elevated plasma TAFI levels have been associated with an increased risk of both DVT and cardiovascular disease.

The metabolic syndrome also is accompanied by altered fibrinolytic activity. This syndrome, which comprises abdominal fat (central obesity), altered glucose and insulin metabolism, dyslipidemia, and hypertension, has been associated with atherothrombosis. The mechanism for enhanced thrombosis appears to be due both to altered platelet function and to a procoagulant and hypofibrinolytic state. One of the most frequently documented prothrombotic abnormalities reported in this syndrome is an increase in plasma levels of PAI-1.

In addition to contributing to platelet function, inflammation plays a role in both coagulation-dependent thrombus formation and thrombus resolution. Both polymorphonuclear neutrophils and monocytes/macrophages contribute to multiple overlapping thrombotic functions, including fibrinolysis, chemokine and cytokine production, and phagocytosis.

THE DISTINCTION BETWEEN ARTERIAL AND VENOUS THROMBOSIS

Although there is overlap, venous thrombosis and arterial thrombosis are initiated differently, and clot formation progresses by somewhat distinct pathways. In the setting of stasis or states of hypercoagulability, venous thrombosis is activated with the initiation of the coagulation cascade primarily due to exposure of tissue factor; this leads to the formation of thrombin and the subsequent conversion of fibrinogen to fibrin. In the artery, thrombin formation also occurs, but thrombosis is primarily promoted by the adhesion of platelets to an injured vessel and stimulated by exposed extracellular matrix (Figs. 117-1 and 117-2). There is wide variation in individual responses to vascular injury, an important determinant of which is the predisposition an individual has to arterial or venous thrombosis. This concept has been supported indirectly in prothrombotic animal models in which there is poor correlation between the propensity to develop venous versus arterial thrombosis.

Despite considerable progress in understanding the role of hypercoagulable states in VTE, the contribution of hypercoagulability to arterial vascular disease is much less well understood. Although specific thrombophilic conditions, such as factor V Leiden and the prothrombin G20210A mutation, are risk factors for DVT, pulmonary embolism, and other venous thromboembolic events, their contribution to arterial thrombosis is less well defined. In fact, to the contrary, many of these thrombophilic factors have not been found to be clinically important risk factors for arterial thrombotic events, such as acute coronary syndromes.

Clinically, although the pathophysiology is distinct, arterial and venous thrombosis do share common risk factors, including age, obesity, cigarette smoking, diabetes mellitus, arterial hypertension, hyperlipidemia, and metabolic syndrome. Select genetic variants, including those of the glutathione peroxidase-3 (GPx3) gene, have also been associated with arterial and venous thrombo-occlusive disease. Importantly, arterial and venous thrombosis may both be triggered by pathophysiologic stimuli responsible for activating inflammatory and oxidative pathways.

The diagnosis and treatment of ischemic heart disease are discussed in [Chap. 273](#). Stroke diagnosis and management are discussed in [Chap. 307](#). The diagnosis and management of DVT and PE are discussed in [Chap. 279](#).

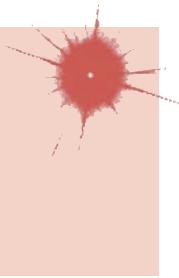
FURTHER READING

- ACKERMANN M et al: Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. *N Engl J Med* 383:120, 2020.
- BARON TH et al: Management of antithrombotic therapy in patients undergoing invasive procedures. *N Engl J Med* 368:2113, 2013.
- BECATTINI C, AEGNELLI G: Acute treatment of venous thromboembolism. *Blood* 135:305, 2020.
- ENGELMANN B, MASSBERG S: Thrombosis as an intravascular effector of innate immunity. *Nat Rev Immunol* 13:34, 2013.
- FURIE B, FURIE BC: Mechanisms of thrombus formation. *N Engl J Med* 359:938, 2008.
- KOUPENOVA M et al: Thrombosis and platelets: An update. *Eur Heart J* 38:785, 2017.
- MACKMAN N: New insights into the mechanisms of venous thrombosis. *J Clin Invest* 122:2331, 2012.
- MOZAFFARIAN D et al: Heart disease and stroke statistics—A report from the American Heart Association 2016 update. *Circulation* 133:447, 2016.
- TAPSON VF: Acute pulmonary embolism. *N Engl J Med* 358:1037, 2008.
- TICHELAAR YI et al: Infections and inflammatory diseases as risk factors for venous thrombosis. A systematic review. *Thromb Haemost* 107:827, 2012.

118

Antiplatelet, Anticoagulant, and Fibrinolytic Drugs

Jeffrey I. Weitz



Thromboembolic disorders are major causes of morbidity and mortality. Thrombosis can occur in arteries or veins. Arterial thrombosis is the most common cause of acute myocardial infarction (MI), ischemic stroke, and limb gangrene. Venous thromboembolism encompasses deep vein thrombosis (DVT), which can lead to postthrombotic syndrome, and pulmonary embolism (PE), which can be fatal or can result in chronic thromboembolic pulmonary hypertension.

Most arterial thrombi are superimposed on disrupted atherosclerotic plaque because plaque rupture exposes thrombogenic material in the core to the blood. This material then triggers platelet aggregation and fibrin formation, which results in the generation of a platelet-rich thrombus that can temporarily or permanently occlude blood flow. In contrast, venous thrombi rarely form at sites of obvious vascular disruption. Although they can develop after surgical trauma to veins or secondary to indwelling venous catheters, venous thrombi usually originate in the valve cusps of the deep veins of the calf or in the muscular sinuses. Sluggish blood flow reduces the oxygen supply to the avascular valve cusps. Endothelial cells lining these valve cusps become activated and express adhesion molecules on their surface. Tissue factor-bearing leukocytes and microvesicles adhere to these activated cells and induce coagulation. DNA extruded from neutrophils forms neutrophil extracellular traps (NETs) that provide a scaffold that binds platelets and promotes their activation and aggregation and activate factor XII. Local thrombus formation is exacerbated by reduced clearance of activated clotting factors because of impaired blood flow. If the thrombi extend from the calf veins into the popliteal and more proximal veins of the leg, thrombus fragments can dislodge, travel to the lungs, and produce a PE.

Arterial and venous thrombi are composed of platelets, fibrin, and trapped red blood cells, but the proportions differ. Arterial thrombi are rich in platelets because of the high shear in the injured arteries. In contrast, venous thrombi, which form under low shear conditions, contain relatively few platelets and are predominantly composed of fibrin and trapped red cells. Because of the predominance of platelets, arterial thrombi appear white, whereas venous thrombi are red in color, reflecting the trapped red cells.

Antithrombotic drugs are used for prevention and treatment of thrombosis. Targeting the components of thrombi, these agents include (1) antiplatelet drugs, (2) anticoagulants, and (3) fibrinolytic agents ([Fig. 118-1](#)). With the predominance of platelets in arterial thrombi, strategies to attenuate arterial thrombosis focus mainly on antiplatelet agents, although, in the acute setting, they may include anticoagulants and fibrinolytic agents. The addition of low-dose rivaroxaban, an oral factor Xa inhibitor, to dual-antiplatelet therapy reduces recurrent ischemic events and stent thrombosis in patients with acute coronary syndrome, whereas its addition to aspirin reduces the risk of major adverse coronary and limb events in patients with stable coronary or peripheral artery disease. These findings highlight the utility of combining low

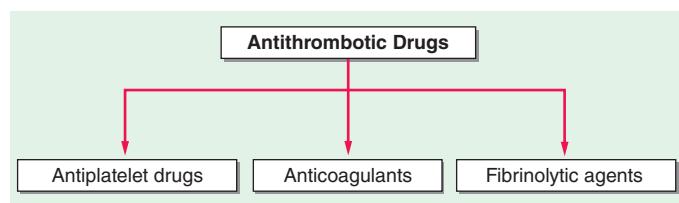


FIGURE 118-1 Classification of antithrombotic drugs.

dose anticoagulants with antiplatelet agents for secondary prevention in patients at risk for recurrent atherothrombotic events.

Anticoagulants are the mainstay of prevention and treatment of venous thromboembolism because fibrin is the predominant component of venous thrombi. Antiplatelet drugs are less effective than anticoagulants in this setting because of the limited platelet content of venous thrombi. Fibrinolytic therapy is used in selected patients with venous thromboembolism. For example, patients with massive PE can benefit from systemic or catheter-directed fibrinolytic therapy. Pharmacomechanical therapy also is used to restore blood flow in patients with extensive DVT involving the iliac and/or femoral veins.

ANTIPLATELET DRUGS

ROLE OF PLATELETS IN ARTERIAL THROMBOSIS

In healthy vasculature, circulating platelets are maintained in an inactive state by nitric oxide (NO) and prostacyclin released by endothelial cells lining the blood vessels. In addition, endothelial cells also express CD39 on their surface, a membrane-associated ecto-adenosine diphosphatase (ADPase) that degrades ADP released from activated platelets. When the vessel wall is damaged, release of these substances is impaired and subendothelial matrix is exposed. Platelets adhere to exposed collagen via $\alpha_2\beta_1$ and glycoprotein (Gp) VI and to von Willebrand factor (VWF) via Gp Ib α and Gp IIb/IIIa ($\alpha_{IIb}\beta_3$)—receptors that are constitutively expressed on the platelet surface. Adherent platelets undergo a change in shape, secrete ADP from their dense granules, and synthesize and release thromboxane A₂. Released ADP and thromboxane A₂, which are platelet agonists, activate ambient platelets and recruit them to the site of vascular injury (Fig. 118-2).

Disruption of the vessel wall also exposes tissue factor-expressing cells to the blood. Tissue factor binds factor VIIa and initiates coagulation. Activated platelets potentiate coagulation by providing a surface that binds clotting factors and supports the assembly of activation complexes that enhance thrombin generation. In addition to converting fibrinogen to fibrin, thrombin serves as a potent platelet agonist and recruits more platelets to the site of vascular injury. Thrombin also amplifies its own generation by feedback activation of factors V, VIII, and XI and solidifies the fibrin network by activating factor XIII, which then cross-links the fibrin strands.

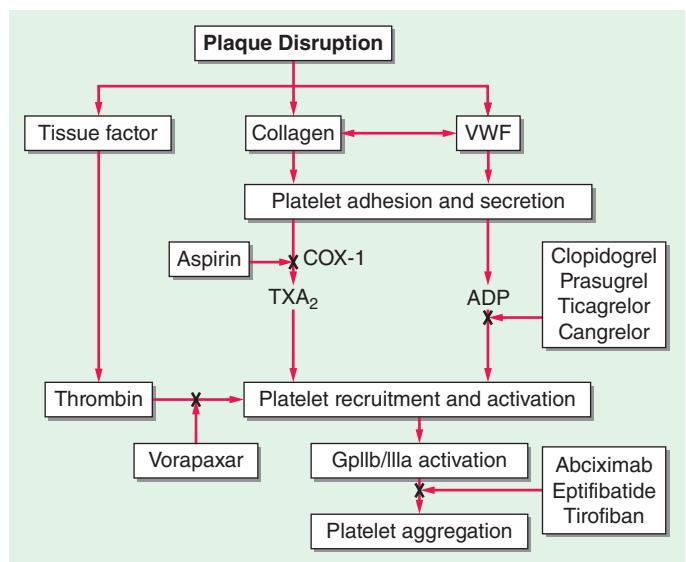


FIGURE 118-3 Site of action of antiplatelet drugs. Aspirin inhibits thromboxane A₂ (TXA₂) synthesis by irreversibly acetylating cyclooxygenase-1 (COX-1). Reduced TXA₂ release attenuates platelet activation and recruitment to the site of vascular injury. Clopidogrel and prasugrel irreversibly block P2Y₁₂, a key ADP receptor on the platelet surface; cangrelor and ticagrelor are reversible inhibitors of P2Y₁₂. Abciximab, eptifibatide, and tirofiban inhibit the final common pathway of platelet aggregation by blocking fibrinogen and von Willebrand factor binding to activated glycoprotein (Gp) IIb/IIIa. Vorapaxar inhibits thrombin-mediated platelet activation by targeting protease-activated receptor-1 (PAR-1), the major thrombin receptor on human platelets.

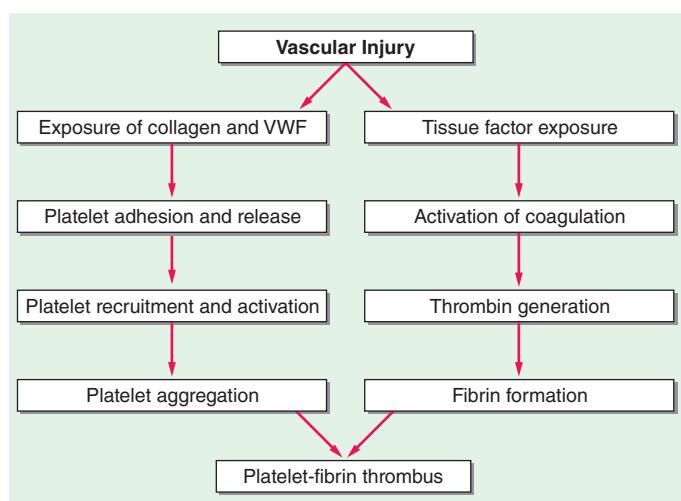


FIGURE 118-2 Coordinated role of platelets and the coagulation system in thrombogenesis. Vascular injury simultaneously triggers platelet activation and aggregation and activation of the coagulation system. Platelet activation is initiated by exposure of subendothelial collagen and von Willebrand factor (VWF), onto which platelets adhere. Adherent platelets become activated and release ADP and thromboxane A₂, platelet agonists that activate ambient platelets and recruit them to the site of injury. When platelets are activated, glycoprotein IIb/IIIa on their surface undergoes a conformational change that enables it to ligate fibrinogen and/or VWF and mediate platelet aggregation. Coagulation is triggered by tissue factor exposed at the site of injury. Tissue factor triggers thrombin generation. As a potent platelet agonist, thrombin amplifies platelet recruitment to the site of injury. Thrombin also converts fibrinogen to fibrin, and the fibrin strands then weave the platelet aggregates together to form a platelet/fibrin thrombus.

When platelets are activated, Gp IIb/IIIa, the most abundant receptor on the platelet surface, undergoes a conformational change that enables it to bind fibrinogen and, under high shear conditions, VWF. Divalent fibrinogen or multivalent VWF molecules bridge adjacent platelets together to form platelet aggregates. Fibrin strands, generated through the action of thrombin, then weave these aggregates together to form a platelet/fibrin mesh.

Antiplatelet drugs target various steps in this process. The commonly used drugs include aspirin, ADP receptor inhibitors, which include the thienopyridines (clopidogrel and prasugrel) as well as ticagrelor and cangrelor, dipyridamole, Gp IIb/IIIa antagonists, and vorapaxar.

ASPIRIN

The most widely used antiplatelet agent worldwide is aspirin. As a cheap and effective antiplatelet drug, aspirin serves as the foundation of most antiplatelet strategies.

Mechanism of Action Aspirin produces its antithrombotic effect by irreversibly acetylating and inhibiting platelet cyclooxygenase (COX)-1 (Fig. 118-3), a critical enzyme in the biosynthesis of thromboxane A₂. At high doses (~1 g/d), aspirin also inhibits COX-2, an inducible COX isoform found in endothelial cells and inflammatory cells. In endothelial cells, COX-2 initiates the synthesis of prostacyclin, a potent vasodilator and inhibitor of platelet aggregation.

Indications Aspirin is widely used for secondary prevention of cardiovascular events in patients with established coronary artery, cerebral artery, or peripheral artery disease. Compared with placebo in this setting, aspirin produces a 25% reduction in the risk of cardiovascular death, MI, or stroke. Use of aspirin for primary prevention is controversial. Recent studies have questioned whether the benefits of daily aspirin for primary cardiac protection outweigh its associated risks for gastrointestinal and intracerebral hemorrhage. Consequently, aspirin is no longer recommended for primary cardiac prevention unless the baseline cardiovascular risk is at least 1% per year and 10% at 10 years and patients are at low risk for bleeding.

Dosages Aspirin is usually administered at doses of 75–325 mg once daily. Higher doses of aspirin are not more effective than lower aspirin doses, and some analyses suggest reduced efficacy with higher doses. Because the side effects of aspirin are dose-related, daily aspirin doses of 75–100 mg are recommended for most indications. When rapid platelet inhibition is required, an initial aspirin dose of at least 160 mg should be given.

Side Effects The most common side effects are gastrointestinal and range from dyspepsia to erosive gastritis or peptic ulcers with bleeding and perforation. These side effects are dose-related. Use of enteric-coated or buffered aspirin in place of plain aspirin does not eliminate gastrointestinal side effects. The overall risk of major bleeding with aspirin is 1–3% per year. The risk of bleeding is increased two- to threefold when aspirin is given in conjunction with other antiplatelet drugs, such as clopidogrel, or with anticoagulants, such as warfarin. When dual or triple therapy is prescribed, low-dose aspirin should be given (75–100 mg daily). Eradication of *Helicobacter pylori* infection and administration of proton pump inhibitors may reduce the risk of aspirin-induced upper gastrointestinal bleeding in patients with peptic ulcer disease.

Aspirin should not be administered to patients with a history of aspirin allergy characterized by bronchospasm. This problem occurs in ~0.3% of the general population but is more common in those with chronic urticaria or asthma, particularly in individuals with nasal polyps or chronic rhinitis. Hepatic and renal toxicity are observed with aspirin overdose.

Aspirin Resistance Clinical aspirin resistance is defined as the failure of aspirin to protect patients from ischemic vascular events. This is not a helpful definition because it is made after the event occurs. Furthermore, it is not realistic to expect aspirin, which only blocks thromboxane A₂-induced platelet activation, to prevent all vascular events.

Aspirin resistance has also been described biochemically as failure of the drug to produce its expected inhibitory effects on tests of platelet function, such as thromboxane A₂ synthesis or arachidonic acid-induced platelet aggregation. Potential causes of aspirin resistance include poor compliance, reduced absorption, drug-drug interaction with ibuprofen, and overexpression of COX-2. Unfortunately, the tests for aspirin resistance have not been well standardized, and there is little evidence that they identify patients at increased risk of recurrent vascular events, or that resistance can be reversed by giving higher doses of aspirin or by adding other antiplatelet drugs. Until such information is available, testing for aspirin resistance remains a research tool.

■ ADP RECEPTOR ANTAGONISTS

The ADP receptor antagonists include the thienopyridines (clopidogrel and prasugrel) as well as ticagrelor and cangrelor. All of these drugs target P2Y₁₂, the key ADP receptor on platelets.

Thienopyridines • MECHANISM OF ACTION The thienopyridines are structurally related drugs that selectively inhibit ADP-induced platelet aggregation by irreversibly blocking P2Y₁₂ (Fig. 118-3). Clopidogrel and prasugrel are prodrugs that require metabolic activation by the hepatic cytochrome P450 (CYP) enzyme system. Prasugrel is about 10-fold more potent than clopidogrel and has a more rapid onset of action because of better absorption and more streamlined metabolic activation.

INDICATIONS When compared with aspirin in patients with recent ischemic stroke, recent MI, or a history of peripheral arterial disease, clopidogrel reduced the risk of cardiovascular death, MI, and stroke by 8.7%. Therefore, clopidogrel is more effective than aspirin but is also more expensive. Clopidogrel and aspirin are often combined to capitalize on their capacity to block complementary pathways of platelet activation. For example, the combination of aspirin plus clopidogrel is recommended for at least 4 weeks after implantation of a bare metal stent in a coronary artery and for at least a year in those with a drug-eluting stent. Concerns about late in-stent thrombosis with drug-eluting stents have led some experts to recommend long-term use of clopidogrel plus aspirin for the latter indication.

The combination of clopidogrel and aspirin is also effective in patients with unstable angina. Thus, in 12,562 such patients, the risk of cardiovascular death, MI, or stroke was 9.3% in those randomized to the combination of clopidogrel and aspirin and 11.4% in those given aspirin alone. This 20% relative risk reduction with combination therapy was highly statistically significant. However, combining clopidogrel with aspirin increases the risk of major bleeding to about 2% per year. This bleeding risk persists even if the daily dose of aspirin is ≤100 mg. Therefore, the combination of clopidogrel and aspirin should only be used when there is a clear benefit. For example, this combination has not proven to be superior to clopidogrel alone in patients with acute ischemic stroke or to aspirin alone for primary prevention in those at risk for cardiovascular events.

Prasugrel was compared with clopidogrel in 13,608 patients with acute coronary syndromes who were scheduled to undergo percutaneous coronary intervention. The incidence of the primary efficacy endpoint, a composite of cardiovascular death, MI, or stroke, was significantly lower with prasugrel than with clopidogrel (9.9% and 12.1%, respectively), mainly reflecting a reduction in the incidence of nonfatal MI. The incidence of stent thrombosis also was significantly lower with prasugrel (1.1% and 2.4%, respectively). However, these advantages were at the expense of significantly higher rates of fatal bleeding (0.4% and 0.1%, respectively) and life-threatening bleeding (1.4% and 0.9%, respectively) with prasugrel. Because patients older than age 75 years and those with a history of prior stroke or transient ischemic attack have a particularly high risk of bleeding, prasugrel should generally be avoided in older patients, and the drug is contraindicated in those with a history of cerebrovascular disease. Caution is required if prasugrel is used in patients weighing less than 60 kg or in those with renal impairment.

When prasugrel was compared with clopidogrel in 7243 patients with unstable angina or MI without ST-segment elevation, prasugrel failed to reduce the rate of the primary efficacy endpoint, which was a composite of cardiovascular death, MI, and stroke. Because of the negative results of this study, prasugrel is reserved for patients undergoing percutaneous coronary intervention. In this setting, prasugrel is usually given in conjunction with aspirin. To reduce the risk of bleeding, the daily aspirin dose should be ≤100 mg.

For patients with noncardioembolic stroke or high-risk transient ischemic attack, the combination of clopidogrel or ticagrelor plus aspirin for 21–30 days followed by aspirin alone thereafter reduces the risk of stroke, MI, and vascular death by up to 30% compared with aspirin alone. Therefore, dual antiplatelet therapy is often administered for the first 3–4 weeks in such patients.

DOSING Clopidogrel is given once daily at a dose of 75 mg. Loading doses of clopidogrel are given when rapid ADP receptor blockade is desired. For example, patients undergoing coronary stenting are often given a loading dose of 300–600 mg, which produces inhibition of ADP-induced platelet aggregation in about 4–6 h. After a loading dose of 60 mg, prasugrel is given once daily at a dose of 10 mg. Patients older than age 75 years or weighing less than 60 kg should receive a lower daily prasugrel dose of 5 mg.

SIDE EFFECTS The most common side effect of clopidogrel and prasugrel is bleeding. Because of its greater potency, bleeding is more common with prasugrel than clopidogrel. To reduce the risk of bleeding, clopidogrel and prasugrel should be stopped 5–7 days before major surgery. In patients taking clopidogrel or prasugrel who present with serious bleeding, platelet transfusion may be helpful.

Hematologic side effects, including neutropenia, thrombocytopenia, and thrombotic thrombocytopenic purpura, are rare.

THIENOPYRIDINE RESISTANCE The capacity of clopidogrel to inhibit ADP-induced platelet aggregation varies among subjects. This variability reflects, at least in part, genetic polymorphisms in the CYP isoenzymes involved in the metabolic activation of clopidogrel. Most important of these is CYP2C19. Clopidogrel-treated patients with the loss-of-function CYP2C19*2 allele exhibit reduced platelet inhibition compared with those with the wild-type CYP2C19*1 allele and experience a higher rate of cardiovascular events. This is important because

estimates suggest that up to 25% of whites, 30% of African Americans, and 50% of Asians carry the loss-of-function allele, which would render them resistant to clopidogrel. Even patients with the reduced function CYP2C19^{*3}, ^{*4}, or ^{*5} alleles may derive less benefit from clopidogrel than those with the full-function CYP2C19^{*1} allele. Concomitant administration of clopidogrel with proton pump inhibitors, which are inhibitors of CYP2C19, produces a small reduction in the inhibitory effects of clopidogrel on ADP-induced platelet aggregation. The extent to which this interaction increases the risk of cardiovascular events remains controversial.

In contrast to their effect on the metabolic activation of clopidogrel, CYP2C19 polymorphisms appear to be less important determinants of the activation of prasugrel. Thus, no association was detected between the loss-of-function allele and decreased platelet inhibition or increased rate of cardiovascular events with prasugrel. The observation that genetic polymorphisms affecting clopidogrel absorption or metabolism influence clinical outcomes raises the possibilities that pharmacogenetic profiling may be useful to identify clopidogrel-resistant patients and that point-of-care assessment of the extent of clopidogrel-induced platelet inhibition may help detect patients at higher risk for subsequent cardiovascular events. Clinical trials designed to evaluate these possibilities have thus far been negative. Although administration of higher doses of clopidogrel can overcome a reduced response to clopidogrel, the clinical benefit of this approach is uncertain. Instead, prasugrel or ticagrelor may be better choices for these patients.

Ticagrelor As an orally active inhibitor of P2Y₁₂, ticagrelor differs from the thienopyridines in that ticagrelor does not require metabolic activation and it produces reversible inhibition of the ADP receptor.

MECHANISM OF ACTION Like the thienopyridines, ticagrelor inhibits P2Y₁₂. Because it does not require metabolic activation, ticagrelor has a more rapid onset and offset of action than clopidogrel, and it produces greater and more predictable inhibition of ADP-induced platelet aggregation than clopidogrel.

INDICATIONS Ticagrelor is indicated for the secondary prevention of atherothrombotic events in patients with an acute coronary syndrome treated medically or with percutaneous coronary intervention (PCI) with or without stent implantation or with coronary artery bypass graft (CABG) surgery. Ticagrelor also is indicated for up to 3 years for secondary prevention in patients with a prior history of MI at least one year ago who are at high risk for atherothrombotic events. For patients with acute coronary syndrome undergoing PCI, guidelines give preference to ticagrelor over clopidogrel. Guidelines give preference to ticagrelor over clopidogrel, particularly in higher risk patients.

DOSING Ticagrelor is initiated with an oral loading dose of 180 mg followed by 90 mg twice daily. The dose does not require adjustment in patients with renal impairment, but the drug should be used with caution in patients with hepatic disease and in those receiving potent inhibitors or inducers of CYP3A4 because ticagrelor is metabolized in the liver via CYP3A4. Ticagrelor is usually administered in conjunction with aspirin; the daily aspirin dose should not exceed 100 mg.

SIDE EFFECTS In addition to bleeding, the most common side effects of ticagrelor are dyspnea, which can occur in up to 15% of patients, and asymptomatic ventricular pauses. The dyspnea, which tends to occur soon after initiating ticagrelor, is usually self-limiting and mild in intensity. The mechanism responsible for this side effect is unknown.

To reduce the risk of bleeding, ticagrelor should be stopped at least 5 days before major surgery. Platelet transfusion is unlikely to be of benefit in patients with ticagrelor-related bleeding or in those requiring urgent surgery because the drug will bind to P2Y₁₂ on the transfused platelets. Bentracimab, an antibody fragment that binds ticagrelor and its metabolite with high affinity and rapidly reverses their platelet inhibitory effects, is under development for ticagrelor reversal prior to urgent surgery or intervention or for patients with serious bleeding.

Cangrelor Cangrelor is a rapidly acting reversible inhibitor of P2Y₁₂ that is administered intravenously. It has an immediate onset of action,

a half-life of 3–5 min, and an offset of action within an hour. Cangrelor is licensed for use in patients undergoing percutaneous coronary intervention and produces rapid ADP receptor blockade in those who have not received pretreatment with clopidogrel, prasugrel, or ticagrelor.

Cangrelor is administered as a 30 µg/kg IV bolus prior to percutaneous coronary intervention followed by an infusion of 4 µg/kg per minute for at least 2 h or for the duration of the procedure, whichever is longer. When transitioning to oral P2Y₁₂ inhibitor therapy, ticagrelor can be given at a loading dose of 180 mg at any time during the cangrelor infusion or immediately after discontinuation. In contrast, loading doses of prasugrel or clopidogrel (60 and 600 mg, respectively) should only be given after cangrelor is stopped because cangrelor blocks the interaction of their active metabolites with P2Y₁₂.

■ DIPYRIDAMOLE

Dipyridamole is a relatively weak antiplatelet agent on its own, but an extended-release formulation of dipyridamole combined with low-dose aspirin, a preparation known as Aggrenox, is sometimes used for secondary prevention in patients with transient ischemic attacks or ischemic stroke.

Mechanism of Action By inhibiting phosphodiesterase, dipyridamole blocks the breakdown of cyclic adenosine monophosphate (AMP). Increased levels of cyclic AMP reduce intracellular calcium and inhibit platelet activation. Dipyridamole also blocks the uptake of adenosine by platelets and other cells. This produces a further increase in local cyclic AMP levels because the platelet adenosine A₂ receptor is coupled to adenylate cyclase (Fig. 118-4).

Indications Dipyridamole plus aspirin was compared with aspirin or dipyridamole alone, or with placebo, in patients with an ischemic stroke or transient ischemic attack. The combination reduced the risk of stroke by 22.1% compared with aspirin and by 24.4% compared with dipyridamole. A second trial compared dipyridamole plus aspirin with aspirin alone for secondary prevention in patients with ischemic stroke. Vascular death, stroke, or MI occurred in 13% of patients given combination therapy and in 16% of those treated with aspirin alone. Another trial randomized 20,332 patients with noncardioembolic ischemic stroke to either Aggrenox or clopidogrel. The primary efficacy endpoint of recurrent stroke occurred in 9.0% of those given Aggrenox and in 8.8% of patients treated with clopidogrel. Although this difference was not statistically significant, the study failed to meet the prespecified margin to claim noninferiority of Aggrenox relative to clopidogrel. These results have dampened enthusiasm for the use of Aggrenox.

Because of its vasodilatory effects and the paucity of data supporting the use of dipyridamole in patients with symptomatic coronary artery disease, Aggrenox should not be used for stroke prevention in such patients. Clopidogrel is a better choice in this setting.

Dosing Aggrenox is given twice daily. Each capsule contains 200 mg of extended-release dipyridamole and 25 mg of aspirin.

Side Effects Because dipyridamole has vasodilatory effects, it must be used with caution in patients with coronary artery disease. Gastrointestinal complaints, headache, facial flushing, dizziness, and hypotension can also occur. These symptoms often subside with continued use of the drug.

■ GP IIb/IIIa RECEPTOR ANTAGONISTS

As a class, parenteral Gp IIb/IIIa receptor antagonists have a niche in patients with acute coronary syndrome. The three agents in this class are abciximab, eptifibatide, and tirofiban.

Mechanism of Action A member of the integrin family of adhesion receptors, Gp IIb/IIIa is found on the surface of platelets and megakaryocytes. With about 80,000 copies per platelet, Gp IIb/IIIa is the most abundant receptor. Consisting of a noncovalently linked heterodimer, Gp IIb/IIIa is inactive on resting platelets. When platelets

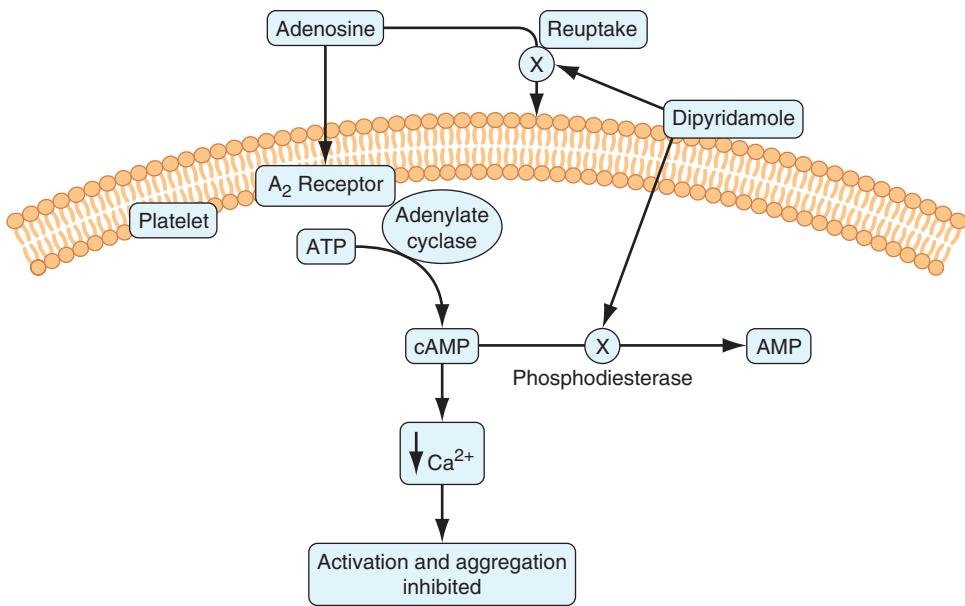


FIGURE 118-4 Mechanism of action of dipyridamole. Dipyridamole increases levels of cyclic AMP (cAMP) in platelets by (1) blocking the reuptake of adenosine and (2) inhibiting phosphodiesterase-mediated cyclic AMP degradation. By promoting calcium uptake, cyclic AMP reduces intracellular levels of calcium. This, in turn, inhibits platelet activation and aggregation.

are activated, inside-outside signal transduction pathways trigger a conformational activation of the receptor. Once activated, Gp IIb/IIIa binds adhesive molecules, such as fibrinogen and, under high shear conditions, VWF. Binding is mediated by the Arg-Gly-Asp (RGD) sequence found on the α chains of fibrinogen and on VWF, and by the Lys-Gly-Asp (KGD) sequence located within a unique dodecapeptide domain on the γ chains of fibrinogen. Once bound, fibrinogen and/or VWF bridge adjacent platelets together to induce platelet aggregation.

Although abciximab, eptifibatide, and tirofiban all target the Gp IIb/IIIa receptor, they are structurally and pharmacologically distinct (Table 118-1). Abciximab is a Fab fragment of a humanized murine monoclonal antibody directed against the activated form of Gp IIb/IIIa. Abciximab binds to the activated receptor with high affinity and blocks the binding of adhesive molecules. In contrast, eptifibatide and tirofiban are synthetic small molecules. Eptifibatide is a cyclic heptapeptide that binds Gp IIb/IIIa because it incorporates the KGD motif, whereas tirofiban is a nonpeptidic tyrosine derivative that acts as an RGD mimetic. Abciximab has a long half-life and can be detected on the surface of platelets for up to 2 weeks; eptifibatide and tirofiban have short half-lives.

Indications Abciximab and eptifibatide are used in patients undergoing percutaneous coronary interventions, particularly those who have not been pretreated with an ADP receptor antagonist. Tirofiban is used in high-risk patients with unstable angina. Eptifibatide also can be used for this indication.

TABLE 118-1 Features of Gp IIb/IIIa Antagonists

FEATURE	ABCIXIMAB	EPTIFIBATIDE	TIROFIBAN
Description	Fab fragment of humanized mouse monoclonal antibody	Cyclical KGD-containing heptapeptide	Nonpeptidic RGD mimetic
Specific for Gp IIb/IIIa	No	Yes	Yes
Plasma half-life	Short (min)	Long (2.5 h)	Long (2.0 h)
Platelet-bound half-life	Long (days)	Short (s)	Short (s)
Renal clearance	No	Yes	Yes

Abbreviation: Gp, glycoprotein.

Dosing All of the Gp IIb/IIIa antagonists are given as an IV bolus followed by an infusion. The recommended dose of abciximab is a bolus of 0.25 mg/kg followed by an infusion of 0.125 µg/kg per minute to a maximum of 10 µg/kg for 12 h. In patients undergoing percutaneous coronary intervention, eptifibatide is given as two 180 µg/kg boluses given 10 min apart, followed by an infusion of 2.0 µg/kg per minute for 18–24 h. For patients with acute coronary syndrome, the second eptifibatide bolus is withheld. Tirofiban is started at a rate of 0.4 µg/kg per minute for 30 min; the drug is then continued at a rate of 0.1 µg/kg per minute for up to 18 h. Because eptifibatide and tirofiban are cleared by the kidneys, the doses must be reduced in patients with renal insufficiency. Thus, the eptifibatide infusion is reduced to 1 µg/kg per minute in patients with a creatinine clearance below 50 mL/min, whereas the dose of tirofiban is cut in half for patients with a creatinine clearance below 30 mL/min.

Side Effects

In addition to bleeding, thrombocytopenia is the most serious complication. Thrombocytopenia is immune-mediated and is caused by antibodies directed against neoantigens on Gp IIb/IIIa that are exposed upon antagonist binding. With abciximab, thrombocytopenia occurs in up to 5% of patients. Thrombocytopenia is severe in ~1% of these individuals. Thrombocytopenia is less common with the other two agents, occurring in ~1% of patients.

VORAPAXAR

An orally active PAR-1 antagonist, vorapaxar blocks thrombin-induced platelet activation. Vorapaxar has a half-life of about 200 h.

Indications When compared with placebo in 12,944 patients with acute coronary syndrome without ST-segment elevation, vorapaxar failed to significantly reduce the primary efficacy endpoint, a composite of cardiovascular death, MI, stroke, recurrent ischemia requiring rehospitalization, and urgent coronary revascularization. Moreover, vorapaxar was associated with increased rates of bleeding, including intracranial bleeding.

In a second trial, vorapaxar was compared with placebo for secondary prevention in 26,449 patients with prior MI, ischemic stroke, or peripheral arterial disease. Overall, vorapaxar reduced the risk for cardiovascular death, MI, or stroke by 13%, but doubled the risk of intracranial bleeding. In the prespecified subgroup of 17,779 patients with prior MI, however, vorapaxar reduced the risk for cardiovascular death, MI, or stroke by 20% compared with placebo (from 9.7% to 8.1%, respectively). The rate of intracranial hemorrhage was higher with vorapaxar than with placebo (0.6% and 0.4%, respectively; $p = .076$) as was the rate of moderate or severe bleeding (3.4% and 2.1%, respectively; $p < .0001$). Based on these data, vorapaxar is licensed for patients younger than 75 years with MI or peripheral artery disease who have no history of stroke, transient ischemic attack, or intracranial bleeding and weigh more than 60 kg.

Dosing Vorapaxar is given at a dose of 2.08 mg once daily.

Side Effects The major side effect is bleeding. Platelet transfusion may be of benefit for vorapaxar reversal.

ANTICOAGULANTS

There are both parenteral and oral anticoagulants. The parenteral anticoagulants include heparin, low-molecular-weight heparin (LMWH), fondaparinux (a synthetic pentasaccharide), lepirudin, desirudin, bivalirudin, and argatroban. Currently available oral anticoagulants

include warfarin; dabigatran etexilate, an oral thrombin inhibitor; and rivaroxaban, apixaban, and edoxaban, which are oral factor Xa inhibitors.

PARENTERAL ANTICOAGULANTS

Heparin A sulfated polysaccharide, heparin is isolated from mammalian tissues rich in mast cells. Most commercial heparin is derived from porcine intestinal mucosa and is a polymer of alternating D-glucuronic acid and N-acetyl-D-glucosamine residues.

MECHANISM OF ACTION Heparin acts as an anticoagulant by activating antithrombin (previously known as antithrombin III) and accelerating the rate at which antithrombin inhibits clotting enzymes, particularly thrombin and factor Xa. Antithrombin, the obligatory plasma cofactor for heparin, is a member of the serine protease inhibitor (serpin) superfamily. Synthesized in the liver and circulating in plasma at a concentration of $2.6 \pm 0.4 \mu\text{M}$, antithrombin acts as a suicide substrate for its target enzymes.

To activate antithrombin, heparin binds to the serpin via a unique pentasaccharide sequence that is found on one-third of the chains of commercial heparin (Fig. 118-5). Heparin chains without this pentasaccharide sequence have little or no anticoagulant activity. Once bound to antithrombin, heparin induces a conformational change in the reactive center loop of antithrombin that renders it more readily accessible to its target proteases. This conformational change enhances the rate at which antithrombin inhibits factor Xa by at least two orders of magnitude but has little effect on the rate of thrombin inhibition. To catalyze thrombin inhibition, heparin serves as a template that binds antithrombin and thrombin simultaneously. Formation of this ternary complex brings the enzyme in close apposition to the inhibitor, thereby promoting the formation of a stable covalent thrombin-antithrombin complex.

Only pentasaccharide-containing heparin chains composed of at least 18 saccharide units (which correspond to a molecular weight of 5400) are of sufficient length to bridge thrombin and antithrombin together. With a mean molecular weight of 15,000, and a range of 5000–30,000, almost all of the chains of unfractionated heparin are long enough to do so. Consequently, by definition, heparin has equal capacity to promote the inhibition of thrombin and factor Xa by antithrombin and is assigned an anti-factor Xa to anti-factor IIa (thrombin) ratio of 1:1.

Heparin causes the release of tissue factor pathway inhibitor (TFPI) from the endothelium. A factor Xa-dependent inhibitor of tissue factor-bound factor VIIa, TFPI may contribute to the antithrombotic activity of heparin. Longer heparin chains induce the release of more TFPI than shorter ones.

PHARMACOLOGY Heparin must be given parenterally. It is usually administered SC or by continuous IV infusion. When used for therapeutic purposes, the IV route is most often employed. If heparin is given SC for treatment of thrombosis, the dose of heparin must be high enough to overcome the limited bioavailability associated with this method of delivery.

In the circulation, heparin binds to the endothelium and to plasma proteins other than antithrombin. Heparin binding to endothelial

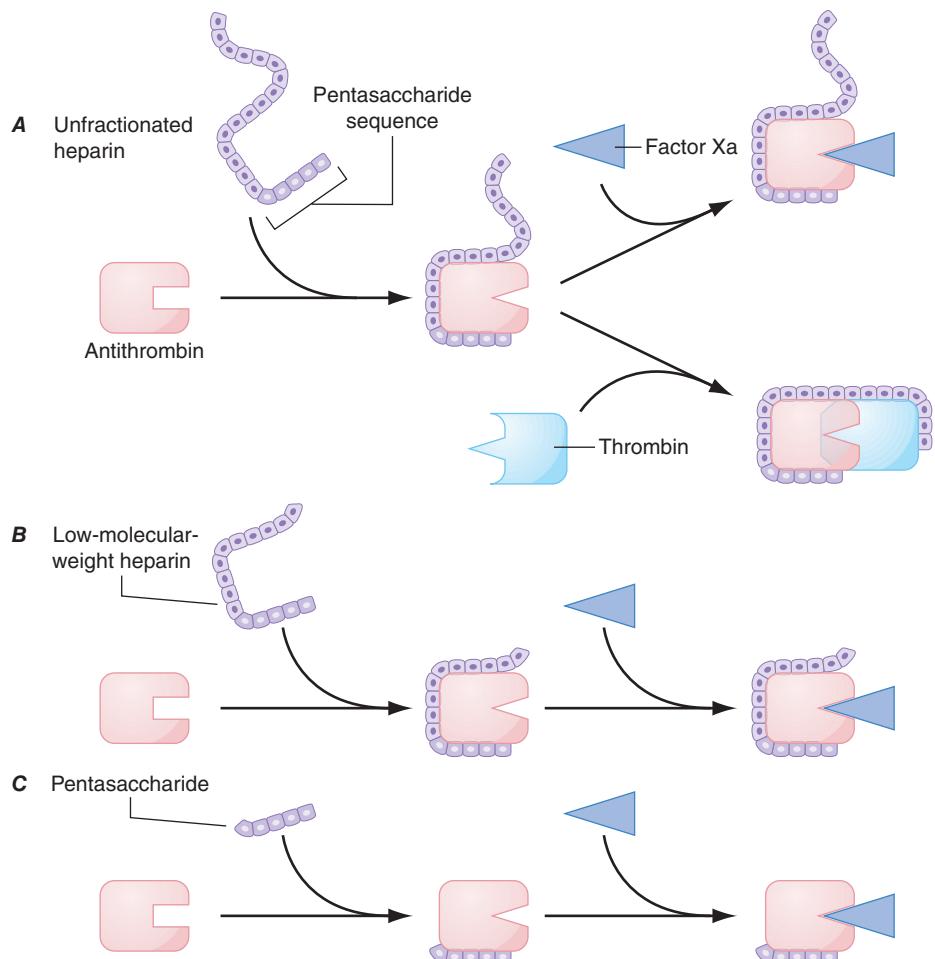


FIGURE 118-5 Mechanism of action of heparin, low-molecular-weight heparin (LMWH), and fondaparinux, a synthetic pentasaccharide. A. Heparin binds to antithrombin via its pentasaccharide sequence. This induces a conformational change in the reactive center loop of antithrombin that accelerates its interaction with factor Xa. To potentiate thrombin inhibition, heparin must simultaneously bind to antithrombin and thrombin. Only heparin chains composed of at least 18 saccharide units, which corresponds to a molecular weight of 5400, are of sufficient length to perform this bridging function. With a mean molecular weight of 15,000, all of the heparin chains are long enough to do this. **B.** LMWH has greater capacity to potentiate factor Xa inhibition by antithrombin than thrombin because, with a mean molecular weight of 4500–5000, at least half of the LMWH chains are too short to bridge antithrombin to thrombin. **C.** The pentasaccharide only accelerates factor Xa inhibition by antithrombin because the pentasaccharide is too short to bridge antithrombin to thrombin.

cells explains its dose-dependent clearance. At low doses, the half-life of heparin is short because it binds rapidly to the endothelium. With higher doses of heparin, the half-life is longer because heparin is cleared more slowly once the endothelium is saturated. Clearance is mainly extra renal; heparin binds to macrophages, which internalize and depolymerize the long heparin chains and secrete shorter chains back into the circulation. Because of its dose-dependent clearance mechanism, the plasma half-life of heparin ranges from 30 to 60 min with bolus IV doses of 25 and 100 units/kg, respectively.

Once heparin enters the circulation, it binds to plasma proteins other than antithrombin, a phenomenon that reduces its anticoagulant activity. Some of the heparin-binding proteins found in plasma are acute-phase reactants whose levels are elevated in ill patients. Others, such as high-molecular-weight multimers of VWF, are released from activated platelets or endothelial cells. Activated platelets also release platelet factor 4 (PF4), a highly cationic protein that binds heparin with high affinity. The large amounts of PF4 found in the vicinity of platelet-rich arterial thrombi can neutralize the anticoagulant activity of heparin. This phenomenon may attenuate heparin's capacity to suppress thrombus growth.

Because the levels of heparin-binding proteins in plasma vary from person to person, the anticoagulant response to fixed or weight-adjusted doses of heparin is unpredictable. Consequently, coagulation

monitoring is essential to ensure that a therapeutic response is obtained. This is particularly important when heparin is administered for treatment of established thrombosis because a subtherapeutic anticoagulant response may render patients at risk for recurrent thrombosis, whereas excessive anticoagulation increases the risk of bleeding.

MONITORING THE ANTICOAGULANT EFFECT Heparin therapy can be monitored using the activated partial thromboplastin time (aPTT) or anti-factor Xa level. Although the aPTT is the test most often used for this purpose, there are problems with this assay. aPTT reagents vary in their sensitivity to heparin, and the type of coagulometer used for testing can influence the results. Consequently, laboratories must establish a therapeutic aPTT range with each reagent-coagulometer combination by measuring the aPTT and anti-factor Xa level in plasma samples collected from heparin-treated patients. For most of the aPTT reagents and coagulometers in current use, therapeutic heparin levels are achieved with a two- to threefold prolongation of the aPTT. Anti-factor Xa levels also can be used to monitor heparin therapy. With this test, therapeutic heparin levels range from 0.3 to 0.7 units/mL.

Up to 25% of heparin-treated patients with venous thromboembolism require >35,000 units/d to achieve a therapeutic aPTT. These patients are considered heparin resistant. It is useful to measure anti-factor Xa levels in heparin-resistant patients because many will have a therapeutic anti-factor Xa level despite a subtherapeutic aPTT. This dissociation in test results occurs because elevated plasma levels of fibrinogen and factor VIII, both of which are acute-phase proteins, shorten the aPTT but have no effect on anti-factor Xa levels. Heparin therapy in patients who exhibit this phenomenon is best monitored using anti-factor Xa levels instead of the aPTT. Patients with congenital or acquired antithrombin deficiency and those with elevated levels of heparin-binding proteins may also need high doses of heparin to achieve a therapeutic aPTT or anti-factor Xa level. If there is good correlation between the aPTT and the anti-factor Xa levels, either test can be used to monitor heparin therapy.

DOSING For prophylaxis, heparin is usually given in fixed doses of 5000 units SC two or three times daily. With these low doses, coagulation monitoring is unnecessary. In contrast, monitoring is essential when the drug is given in therapeutic doses. Fixed-dose or weight-based heparin nomograms are used to standardize heparin dosing and to shorten the time required to achieve a therapeutic anticoagulant response. At least two heparin nomograms have been validated in patients with venous thromboembolism and reduce the time required to achieve a therapeutic aPTT. Weight-adjusted heparin nomograms have also been evaluated in patients with acute coronary syndromes. After an IV heparin bolus of 5000 units or 70 units/kg, a heparin infusion rate of 12–15 units/kg per hour is usually administered. In contrast, weight-adjusted heparin nomograms for patients with venous thromboembolism use an initial bolus of 5000 units or 80 units/kg, followed by an infusion of 18 units/kg per hour. Thus, patients with venous thromboembolism appear to require higher doses of heparin to achieve a therapeutic aPTT than do patients with acute coronary syndromes. This may reflect differences in the thrombus burden. Heparin binds to fibrin, and the amount of fibrin in patients with extensive DVT is greater than that in those with coronary thrombosis.

LIMITATIONS Heparin has pharmacokinetic and biophysical limitations (**Table 118-2**). The pharmacokinetic limitations reflect heparin's propensity to bind in a pentasaccharide-independent fashion to cells and plasma proteins. Heparin binding to endothelial cells explains its dose-dependent clearance, whereas binding to plasma proteins results in a variable anticoagulant response and can lead to heparin resistance.

The biophysical limitations of heparin reflect the inability of the heparin-antithrombin complex to inhibit factor Xa when it is incorporated into the prothrombinase complex, the complex that converts prothrombin to thrombin, and to inhibit thrombin bound to fibrin. Consequently, factor Xa bound to activated platelets within platelet-rich thrombi has the potential to generate thrombin, even in the face of heparin. Once this thrombin binds to fibrin, it too is protected

TABLE 118-2 Pharmacokinetic and Biophysical Limitations of Heparin

LIMITATIONS	MECHANISM
Poor bioavailability at low doses	Binds to endothelial cells and macrophages
Dose-dependent clearance	Binds to macrophages
Variable anticoagulant response	Binds to plasma proteins whose levels vary from patient to patient
Reduced activity in the vicinity of platelet-rich thrombi	Neutralized by platelet factor 4 released from activated platelets
Limited activity against factor Xa incorporated in the prothrombinase complex and thrombin bound to fibrin	Reduced capacity of heparin-antithrombin complex to inhibit factor Xa bound to activated platelets and thrombin bound to fibrin

from inhibition by the heparin-antithrombin complex. Clot-associated thrombin can then trigger thrombus growth by locally activating platelets and amplifying its own generation through feedback activation of factors V, VIII, and XI. Further compounding the problem is the potential for heparin neutralization by the high concentrations of PF4 released from activated platelets within the platelet-rich thrombus.

SIDE EFFECTS The most common side effect of heparin is bleeding. Other complications include thrombocytopenia, osteoporosis, and elevated levels of transaminases.

Bleeding The risk of bleeding rises as the dose of heparin is increased. Concomitant administration of drugs that affect hemostasis, such as antiplatelet or fibrinolytic agents, increases the risk of bleeding, as does recent surgery or trauma. Heparin-treated patients with serious bleeding can be given protamine sulfate to neutralize the heparin. Protamine sulfate, a mixture of basic polypeptides isolated from salmon sperm, binds heparin with high affinity, and the resultant protamine-heparin complexes are then cleared. Typically, 1 mg of protamine sulfate neutralizes 100 units of heparin. Protamine sulfate is given IV. Anaphylactoid reactions to protamine sulfate can occur, and drug administration by slow IV infusion is recommended to reduce the risk.

Thrombocytopenia Heparin can cause thrombocytopenia. Heparin-induced thrombocytopenia (HIT) is an antibody-mediated process that is triggered by antibodies directed against neoantigens on PF4 that are exposed when heparin binds to this protein. These antibodies, which are usually of the IgG isotype, bind simultaneously to the heparin-PF4 complex and to platelet Fc receptors. Such binding activates the platelets and generates platelet microparticles. Circulating microparticles are prothrombotic because they express anionic phospholipids on their surface and can bind clotting factors and promote thrombin generation.

The clinical features of HIT are illustrated in **Table 118-3**. Typically, HIT occurs 5–14 days after initiation of heparin therapy, but it can manifest earlier if the patient has received heparin within the past 3 months. A platelet count <100,000/ μ L or a 50% decrease in the platelet count from the pretreatment value should raise the suspicion of HIT. HIT is more common in surgical patients than in medical patients and, like many autoimmune disorders, occurs more frequently in females than in males.

HIT can be associated with thrombosis, either arterial or venous. Venous thrombosis, which manifests as DVT and/or PE, is more

TABLE 118-3 Features of Heparin-Induced Thrombocytopenia

FEATURES	DETAILS
Thrombocytopenia	Platelet count of $\leq 100,000/\mu\text{L}$ or a decrease in platelet count of $\geq 50\%$
Timing	Platelet count falls 5–14 days after starting heparin
Type of heparin	More common with unfractionated heparin than low-molecular-weight heparin
Type of patient	More common in surgical patients and patients with cancer than general medical patients; more common in women than in men
Thrombosis	Venous thrombosis more common than arterial thrombosis

TABLE 118-4 Management of Heparin-Induced Thrombocytopenia

Stop all heparin.
Give an alternative anticoagulant, such as argatroban, bivalirudin, fondaparinux, or rivaroxaban.
Do not give platelet transfusions.
Do not give warfarin until the platelet count returns to its baseline level. If warfarin was administered, give vitamin K to restore the INR to normal.
Evaluate for thrombosis, particularly deep vein thrombosis.

Abbreviation: INR, international normalized ratio.

common than arterial thrombosis. Arterial thrombosis can manifest as ischemic stroke or acute MI. Rarely, platelet-rich thrombi in the distal aorta or iliac arteries can cause critical limb ischemia.

The diagnosis of HIT is established using enzyme-linked assays to detect antibodies against heparin-PF4 complexes or with platelet activation assays. Enzyme-linked assays are sensitive but can be positive in the absence of any clinical evidence of HIT. The most specific diagnostic test for HIT is the serotonin release assay. This test is performed by quantifying serotonin release when washed platelets loaded with labeled serotonin are exposed to patient serum in the absence or presence of varying concentrations of heparin. If the patient serum contains the HIT antibody, heparin addition induces platelet activation and serotonin release.

Management of HIT is outlined in **Table 118-4**. Heparin should be stopped in patients with suspected or documented HIT, and an alternative anticoagulant should be administered to prevent or treat thrombosis. The agents most often used for this indication are parenteral direct thrombin inhibitors, such as argatroban or bivalirudin, or factor Xa inhibitors, such as fondaparinux or rivaroxaban. A HIT-like syndrome known as vaccine induced thrombotic thrombocytopenia is a rare complication after vaccination with adenovirus COVID-19 vaccines. Characterized by thrombosis and thrombocytopenia that occur 4 to 28 days after vaccination, patients can present with cerebral or splanchnic vein thrombosis as well as DVT or PE. The diagnosis is established by evidence of antibodies against PF4 and a positive serotonin release assay with added PF4. Treatment can include intravenous immunoglobulin, steroids, and plasma exchange to offset the effects of the antibodies against PF4 and anticoagulants such as argatroban, fondaparinux or rivaroxaban to treat the thrombosis.

Patients with HIT, particularly those with associated thrombosis, often have evidence of increased thrombin generation that can lead to consumption of protein C. If these patients are given warfarin without a concomitant anticoagulant that inhibits thrombin or thrombin generation, the further decrease in protein C levels induced by the vitamin K antagonist can trigger skin necrosis. To avoid this problem, patients with HIT should be treated with a direct thrombin inhibitor or with fondaparinux until the platelet count returns to normal levels. At this point, low-dose warfarin therapy can be introduced, and the parenteral anticoagulant can be discontinued when the international normalized ratio (INR) has been therapeutic for at least 2 days. Alternatively, a direct oral anticoagulant can be given.

Osteoporosis Treatment with therapeutic doses of heparin for >1 month can cause a reduction in bone density. This complication has been reported in up to 30% of patients given long-term heparin therapy, and symptomatic vertebral fractures occur in 2–3% of these individuals.

Heparin causes bone loss both by decreasing bone formation and by enhancing bone resorption. Thus, heparin affects the activity of both osteoblasts and osteoclasts.

Elevated Levels of Transaminases Therapeutic doses of heparin are frequently associated with modest elevations in the serum levels of hepatic transaminases without a concomitant increase in the level of bilirubin. The levels of transaminases rapidly return to normal when the drug is stopped. The mechanism responsible for this phenomenon is unknown.

Low-Molecular-Weight Heparin Consisting of smaller fragments of heparin, LMWH is prepared from unfractionated heparin by

TABLE 118-5 Advantages of LMWH Over Heparin

ADVANTAGE	CONSEQUENCE
Better bioavailability and longer half-life after subcutaneous injection	Can be given subcutaneously once or twice daily for both prophylaxis and treatment
Dose-independent clearance	Simplified dosing
Predictable anticoagulant response	Coagulation monitoring is unnecessary in most patients
Lower risk of heparin-induced thrombocytopenia	Safer than heparin for short- or long-term administration
Lower risk of osteoporosis	Safer than heparin for extended administration

Abbreviation: LMWH, low-molecular-weight heparin.

controlled enzymatic or chemical depolymerization. The mean molecular weight of LMWH is about 5000, one-third the mean molecular weight of unfractionated heparin. LMWH has advantages over heparin (**Table 118-5**) and has replaced heparin for most indications.

MECHANISM OF ACTION Like heparin, LMWH exerts its anticoagulant activity by activating antithrombin. With a mean molecular weight of 5000, which corresponds to about 17 saccharide units, at least half of the pentasaccharide-containing chains of LMWH are too short to bridge thrombin to antithrombin (Fig. 118-5). However, these chains retain the capacity to accelerate factor Xa inhibition by antithrombin because this activity is largely the result of the conformational changes in antithrombin evoked by pentasaccharide binding. Consequently, LMWH catalyzes factor Xa inhibition by antithrombin more than thrombin inhibition. Depending on their unique molecular weight distributions, LMWH preparations have anti-factor Xa to anti-factor IIa ratios ranging from 2:1 to 4:1.

PHARMACOLOGY Although usually given SC, LMWH also can be administered IV if a rapid anticoagulant response is needed. LMWH has pharmacokinetic advantages over heparin. These advantages reflect the fact that shorter heparin chains bind less avidly to endothelial cells, macrophages, and heparin-binding plasma proteins. Reduced binding to endothelial cells and macrophages eliminates the rapid, dose-dependent, and saturable mechanism of clearance that is a characteristic of unfractionated heparin. Instead, the clearance of LMWH is dose-independent and its plasma half-life is longer. Based on measurement of anti-factor Xa levels, LMWH has a plasma half-life of ~4 h. LMWH is cleared almost exclusively by the kidneys, and the drug can accumulate in patients with renal insufficiency.

LMWH exhibits about 90% bioavailability after SC injection. Because LMWH binds less avidly to heparin-binding proteins in plasma than heparin, LMWH produces a more predictable dose response, and resistance to LMWH is rare. With a longer half-life and more predictable anticoagulant response, LMWH can be given SC once or twice daily without coagulation monitoring, even when the drug is given in treatment doses. These properties render LMWH more convenient than unfractionated heparin. Capitalizing on this feature, studies in patients with venous thromboembolism have shown that home treatment with LMWH is as effective and safe as in-hospital treatment with continuous IV infusions of heparin. Outpatient treatment with LMWH streamlines care, reduces health care costs, and increases patient satisfaction.

MONITORING In the majority of patients, LMWH does not require coagulation monitoring. If monitoring is necessary, anti-factor Xa levels must be measured because most LMWH preparations have little effect on the aPTT. Therapeutic anti-factor Xa levels once daily and twice daily doses of LMWH range from 0.5 to 1.2 units/mL and 1.0 to 2.0 units/mL, respectively, when measured 3–4 h after drug administration. When LMWH is given in prophylactic doses, peak anti-factor Xa levels of 0.2–0.5 units/mL are desirable.

Indications for LMWH monitoring include renal impairment and obesity. LMWH monitoring in patients with a creatinine clearance of ≤30 mL/min is advisable to ensure that there is no drug accumulation.

Although weight-adjusted LMWH dosing appears to produce therapeutic anti-factor Xa levels in patients who are overweight, this approach has not been extensively evaluated in those with morbid obesity. It may also be advisable to monitor the anticoagulant activity of LMWH during pregnancy because dose requirements can change, particularly in the third trimester. Monitoring should also be considered in high-risk settings, such as in pregnant women with mechanical heart valves who are given LMWH for prevention of valve thrombosis, and when LMWH is used in treatment doses in infants or children.

DOSING The doses of LMWH recommended for prophylaxis or treatment vary depending on the LMWH preparation. For prophylaxis, once-daily SC doses of 4000–5000 units are often used, whereas doses of 2500–3000 units are given when the drug is administered twice daily. For treatment of venous thromboembolism, a dose of 150–200 units/kg is given if the drug is administered once daily. If a twice-daily regimen is used, a dose of 100 units/kg is given. In patients with unstable angina, LMWH is given SC on a twice-daily basis at a dose of 100–120 units/kg.

SIDE EFFECTS The major complication of LMWH is bleeding. Meta-analyses suggest that the risk of major bleeding is lower with LMWH than with unfractionated heparin. HIT and osteoporosis are less common with LMWH than with unfractionated heparin.

Bleeding Like the situation with heparin, bleeding with LMWH is more common in patients receiving concomitant therapy with anti-platelet or fibrinolytic drugs. Recent surgery, trauma, or underlying hemostatic defects also increase the risk of bleeding with LMWH.

Although protamine sulfate can be used as an antidote for LMWH, protamine sulfate incompletely neutralizes the anticoagulant activity of LMWH because it only binds the longer chains of LMWH. Because longer chains are responsible for catalysis of thrombin inhibition by antithrombin, protamine sulfate completely reverses the anti-factor IIa activity of LMWH. In contrast, protamine sulfate only partially reverses the anti-factor Xa activity of LMWH because the shorter pentasaccharide-containing chains of LMWH do not bind to protamine sulfate. Consequently, patients at high risk for bleeding may be more safely treated with continuous IV unfractionated heparin than with SC LMWH.

Thrombocytopenia The risk of HIT is about fivefold lower with LMWH than with heparin. LMWH binds less avidly to platelets and causes less PF4 release. Furthermore, with lower affinity for PF4 than heparin, LMWH is less likely to induce the conformational changes in PF4 that trigger the formation of HIT antibodies.

LMWH should not be used to treat HIT patients because most HIT antibodies exhibit cross-reactivity with LMWH. This in vitro cross-reactivity is not simply a laboratory phenomenon because there are case reports of thrombosis when HIT patients were switched from heparin to LMWH.

Osteoporosis Because the risk of osteoporosis is lower with LMWH than with heparin, LMWH is a better choice for extended treatment.

Fondaparinux A synthetic analogue of the antithrombin-binding pentasaccharide sequence, fondaparinux differs from LMWH in several ways (**Table 118-6**). Fondaparinux is licensed for thromboprophylaxis

in general medical or surgical patients and in high-risk orthopedic patients and as an alternative to heparin or LMWH for initial treatment of patients with established venous thromboembolism. Although fondaparinux is used in Europe as an alternative to heparin or LMWH in patients with acute coronary syndrome, the drug is not licensed for this indication in the United States.

MECHANISM OF ACTION As a synthetic analogue of the antithrombin-binding pentasaccharide sequence found in heparin and LMWH, fondaparinux has a molecular weight of 1728. Fondaparinux binds only to antithrombin (Fig. 118-5) and is too short to bridge thrombin to antithrombin. Consequently, fondaparinux catalyzes factor Xa inhibition by antithrombin and does not enhance the rate of thrombin inhibition.

PHARMACOLOGY Fondaparinux exhibits complete bioavailability after SC injection. With no binding to endothelial cells or plasma proteins, the clearance of fondaparinux is dose independent, and its plasma half-life is 17 h. The drug is given SC once daily. Because fondaparinux is cleared unchanged via the kidneys, it is contraindicated in patients with a creatinine clearance <30 mL/min and should be used with caution in those with a creatinine clearance <50 mL/min.

Dosing Fondaparinux produces a predictable anticoagulant response after administration in fixed doses because it does not bind to plasma proteins. The drug is given at a dose of 2.5 mg once daily for prevention of venous thromboembolism. For initial treatment of established venous thromboembolism, fondaparinux is given at a dose of 7.5 mg once daily. The dose can be reduced to 5 mg once daily for those weighing <50 kg and increased to 10 mg for those >100 kg. When given in these doses, fondaparinux is as effective as heparin or LMWH for initial treatment of patients with DVT or PE and produces similar rates of bleeding.

Fondaparinux is used at a dose of 2.5 mg once daily in patients with acute coronary syndrome. When this prophylactic dose of fondaparinux was compared with treatment doses of enoxaparin in patients with non-ST-segment elevation acute coronary syndrome, there was no difference in the rate of cardiovascular death, MI, or stroke at 9 days. However, the rate of major bleeding was 50% lower with fondaparinux than with enoxaparin, a difference that likely reflects the fact that the dose of fondaparinux was lower than that of enoxaparin. In acute coronary syndrome patients who require percutaneous coronary intervention, there is a risk of catheter thrombosis with fondaparinux unless adjunctive heparin is given at the time of the procedure.

SIDE EFFECTS Fondaparinux does not cause HIT because it does not bind to PF4. In contrast to LMWH, there is no cross-reactivity of fondaparinux with HIT antibodies. Consequently, fondaparinux appears to be effective for treatment of HIT patients, although large clinical trials supporting its use are lacking.

The major side effect of fondaparinux is bleeding. Fondaparinux has no antidote. Protamine sulfate has no effect on the anticoagulant activity of fondaparinux because it fails to bind to the drug. Recombinant activated factor VII reverses the anticoagulant effects of fondaparinux in volunteers, but it is unknown whether this agent controls fondaparinux-induced bleeding.

PARENTERAL DIRECT THROMBIN INHIBITORS Direct thrombin inhibitors bind directly to thrombin and block its interaction with its substrates. Approved parenteral direct thrombin inhibitors include recombinant hirudins (lepirudin and desirudin), argatroban, and bivalirudin (**Table 118-7**). Lepirudin and desirudin are no longer available. Argatroban is licensed for treatment of patients with HIT, and bivalirudin is approved as an alternative to heparin in patients undergoing percutaneous coronary intervention, including those with HIT.

ARGATROBAN A univalent inhibitor that targets the active site of thrombin, argatroban is metabolized in the liver. Consequently, this drug must be used with caution in patients with hepatic insufficiency. Argatroban is not cleared via the kidneys, so this drug is safer than fondaparinux for HIT patients with renal impairment.

TABLE 118-6 Comparison of LMWH and Fondaparinux

FEATURES	LMWH	FONDAPARINUX
Number of saccharide units	15–17	5
Catalysis of factor Xa inhibition	Yes	Yes
Catalysis of thrombin inhibition	Yes	No
Bioavailability after subcutaneous administration (%)	90	100
Plasma half-life (h)	4	17
Renal excretion	Yes	Yes
Induces release of tissue factor pathway inhibitor	Yes	No
Neutralized by protamine sulfate	Partially	No

TABLE 118-7 Comparison of the Properties of Lepirudin, Bivalirudin, and Argatroban

	LEPIRUDIN/ DESIRUDIN	BIVALIRUDIN	ARGATROBAN
Molecular mass	7000	1980	527
Site(s) of interaction with thrombin	Active site and exosite 1	Active site and exosite 1	Active site
Renal clearance	Yes	No	No
Hepatic metabolism	No	No	Yes
Plasma half-life (min)	60 (IV) 120–180 (SC)	25	45

Argatroban is administered by continuous IV infusion and has a plasma half-life of ~45 min. The aPTT is used to monitor its anticoagulant effect, and the dose is adjusted to achieve an aPTT 1.5–3 times the baseline value, but not to exceed 100 s. Argatroban also prolongs the INR, a feature that can complicate the transitioning of patients to warfarin. This problem can be circumvented by using the levels of factor X to monitor warfarin instead of the INR. Alternatively, argatroban can be stopped for 2–3 h before INR determination.

BIVALIRUDIN A synthetic 20-amino-acid analogue of hirudin, bivalirudin is a divalent thrombin inhibitor. Thus, the N-terminus of bivalirudin interacts with the active site of thrombin, whereas its C-terminus binds to exosite 1. Bivalirudin has a plasma half-life of 25 min, the shortest half-life of all the parenteral direct thrombin inhibitors. Bivalirudin is degraded by peptidases and is partially excreted via the kidneys. When given in high doses in the cardiac catheterization laboratory, the anticoagulant activity of bivalirudin is monitored using the activated clotting time. With lower doses, its activity can be assessed using the aPTT.

Bivalirudin is licensed as an alternative to heparin in patients undergoing percutaneous coronary intervention. Bivalirudin also has been used successfully in HIT patients who require percutaneous coronary intervention or cardiac bypass surgery.

■ ORAL ANTICOAGULANTS

For many years, vitamin K antagonists such as warfarin were the only available oral anticoagulants. This situation changed with the introduction of the direct oral anticoagulants, which include dabigatran, rivaroxaban, apixaban, and edoxaban.

Warfarin A water-soluble vitamin K antagonist initially developed as a rodenticide, warfarin is the coumarin derivative most often prescribed in North America. Like other vitamin K antagonists, warfarin interferes with the synthesis of the vitamin K-dependent clotting proteins, which include prothrombin (factor II) and factors VII, IX, and X. The synthesis of the vitamin K-dependent anticoagulant proteins, proteins C and S, is also reduced by vitamin K antagonists.

MECHANISM OF ACTION All of the vitamin K-dependent clotting factors possess glutamic acid residues at their N termini. A posttranslational modification adds a carboxyl group to the γ -carbon of these residues to generate γ -carboxyglutamic acid. This modification is essential for expression of the activity of these clotting factors because it permits their calcium-dependent binding to negatively charged phospholipid surfaces. The γ -carboxylation process is catalyzed by a vitamin K-dependent carboxylase. Thus, vitamin K from the diet is reduced to vitamin K hydroquinone by vitamin K reductase (Fig. 118-6). Vitamin K hydroquinone serves as a cofactor for the carboxylase enzyme, which in the presence of carbon dioxide replaces the hydrogen on the γ -carbon of glutamic acid residues with a carboxyl group. During this process, vitamin K hydroquinone is oxidized to vitamin K epoxide, which is then reduced to vitamin K by vitamin K epoxide reductase.

Warfarin inhibits vitamin K epoxide reductase (VKOR), thereby blocking the γ -carboxylation process. This results in the synthesis of vitamin K-dependent clotting proteins that are only partially γ -carboxylated. Warfarin acts as an anticoagulant because these

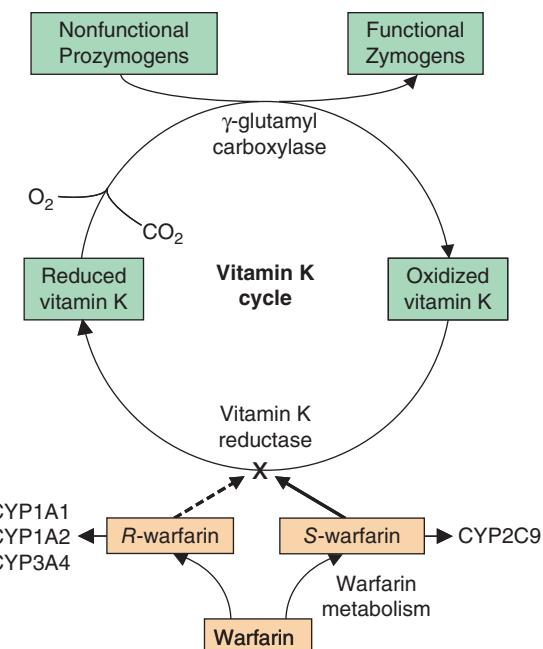


FIGURE 118-6 Mechanism of action of warfarin. A racemic mixture of S- and R-enantiomers, S-warfarin is most active. By blocking vitamin K epoxide reductase, warfarin inhibits the conversion of oxidized vitamin K into its reduced form. This inhibits vitamin K-dependent γ -carboxylation of factors II, VII, IX, and X because reduced vitamin K serves as a cofactor for a γ -glutamyl carboxylase that catalyzes the γ -carboxylation process, thereby converting prozymogens to zymogens capable of binding calcium and interacting with anionic phospholipid surfaces. S-warfarin is metabolized by CYP2C9. Common genetic polymorphisms in this enzyme can influence warfarin metabolism. Polymorphisms in the C1 subunit of vitamin K reductase (*VKORC1*) also can affect the susceptibility of the enzyme to warfarin-induced inhibition, thereby influencing warfarin dosage requirements.

partially γ -carboxylated proteins have little or no biological activity. The onset of action of warfarin is delayed until the newly synthesized clotting factors with reduced activity gradually replace their fully active counterparts.

The antithrombotic effect of warfarin depends on a reduction in the functional levels of factor X and prothrombin, clotting factors that have half-lives of 24 and 72 h, respectively. Because the antithrombotic effect of warfarin is delayed, patients with established thrombosis or at high risk for thrombosis require concomitant treatment with a rapidly acting parenteral anticoagulant, such as heparin, LMWH, or fondaparinux, for at least 5 days.

PHARMACOLOGY Warfarin is a racemic mixture of R and S isomers. Warfarin is rapidly and almost completely absorbed from the gastrointestinal tract. Levels of warfarin in the blood peak about 90 min after drug administration. Racemic warfarin has a plasma half-life of 36–42 h, and >97% of circulating warfarin is bound to albumin. Only the small fraction of unbound warfarin is biologically active.

Warfarin accumulates in the liver where the two isomers are metabolized via distinct pathways. CYP2C9 mediates oxidative metabolism of the more active S isomer (Fig. 118-6). Two relatively common variants, CYP2C9*2 and CYP2C9*3, encode an enzyme with reduced activity. Patients with these variants require lower maintenance doses of warfarin. Approximately 25% of Caucasians have at least one variant allele of CYP2C9*2 or CYP2C9*3, whereas those variant alleles are less common in African Americans and Asians (Table 118-8). Heterozygosity for CYP2C9*2 or CYP2C9*3 decreases the warfarin dose requirement by 20–30% relative to that required in subjects with the wild-type CYP2C9*1/*1 alleles, whereas homozygosity for the CYP2C9*2 or CYP2C9*3 alleles reduces the warfarin dose requirement by 50–70%.

Consistent with their decreased warfarin dose requirement, subjects with at least one CYP2C9 variant allele are at increased risk for bleeding. Compared with individuals with no variant alleles, the risk of warfarin-associated bleeding is almost 2-fold higher in CYP2C9*2 or CYP2C9*3 carriers.